

A11102 436061

NAT'L INST OF STANDARDS & TECH R.I.C.



A11102436061

Howett, Gerald Leona/Linear opponent-col
QC100 .U56 NO.85-3202 V1986 C.1 NBS-PUB-

Linear Opponent-Colors Model Optimized for Brightness Prediction

Reference

NBS

PUBLICATIONS

Gerald L. Howett

U.S. DEPARTMENT OF COMMERCE
National Bureau of Standards
Gaithersburg, MD 20899

June 1985

Issued February 1986

QC

100

.U56

85-3202

1986

sored by

Department of Energy

n, DC

NBSIR 85-3202

**LINEAR OPPONENT-COLORS MODEL
OPTIMIZED FOR BRIGHTNESS
PREDICTION**

Gerald L. Howett

U.S. DEPARTMENT OF COMMERCE
National Bureau of Standards
Gaithersburg, MD 20899

June 1985

Issued February 1986

Partially sponsored by
The U.S. Department of Energy
Washington, DC



U.S. DEPARTMENT OF COMMERCE, Malcolm Baldrige, *Secretary*
NATIONAL BUREAU OF STANDARDS, Ernest Ambler, *Director*

ABSTRACT

Formal multivariate optimization techniques were applied in an attempt to determine how well a linear, opponent-colors model of color vision could account for specific brightness-matching data. The data fitted were from a single experiment by Sanders and Wyszecki that matched an adjustable white light in brightness to each of a set of lights of 96 different colors and constant luminance. A generalized, linear, opponent-colors model was formulated, which included the models of Guth (and coworkers), Ingling (and coworkers), and Thornton as special cases. The model contained 10 parameters, including nine determining the spectral responses of the three opponent-level channels, and one determining the rule for combining the outputs of the three channels to obtain an estimate of equivalent luminance (the luminance of an equally bright white light). Despite difficulties with the optimization procedure, a model was found that correlates better than 0.98 with the fitted data. The predictions of this model for various other color-vision functions were explored and compared with corresponding predictions of the Guth and Lodge model and the Thornton model. The new model's predictions of these functions are less than perfect, but surprisingly good considering that the model was optimized entirely on brightness data (the only restriction being that the luminance channel should not have any negative values). The model was shown to predict the sort of complex mixture of sub- and superadditivity that is present in actual data. Some new algebraic results concerning the "B/Y" or "B/L" (equivalent luminance over luminance) ovals on the chromaticity diagram were derived.

Key words: additivity; brightness; color; equivalent luminance; Guth model; luminance; model, color vision; opponent colors; optimization, multivariate; primaries; white point.

FOREWORD

This report is one of a series documenting National Bureau of Standards research and analysis efforts in support of the Department of Energy/Lawrence Berkeley Laboratory/National Bureau of Standards "Lighting Technology" Program. The research described in this paper was supported partly by the U.S. Department of Energy, through the lighting program of Lawrence Berkeley Laboratory, under Interagency Agreement No. DE-AC03-76SF-00098; and also by the NBS-funded project "Optimized Model for Brightness."

The author acknowledges with thanks the stimulating critiques during the course of the work provided by the Lawrence Berkeley Laboratory Technical Monitors, Drs. Sam Berman and Robert Clear.

An oral report of portions of this study was presented to the Optical Society of America at its fall 1985 annual meeting in Washington, D.C.

ACKNOWLEDGMENTS

The author is grateful to a number of members of the NBS Center for Applied Mathematics for consultations regarding various mathematical aspects of the work described in this report. These generous colleagues are in no way responsible for any statements concerning mathematical matters, or any equations, contained within the report.

In addition, particular thanks are due to Mrs. Donna Shoemaker, who managed to do a difficult typing job very well, under difficult circumstances.

TABLE OF CONTENTS

	<u>Page</u>
ABSTRACT	iii
FOREWORD	iv
ACKNOWLEDGEMENTS	iv
LIST OF TABLES	viii
LIST OF FIGURES	ix
A NOTE ON NOTATION: UNDERLINING	xi
1. INTRODUCTION	1
1.1 <u>HISTORICAL BACKGROUND</u>	2
2. THE NUMERICAL MODELS	4
2.1 <u>GUTH'S MODEL</u>	4
2.2 <u>THORNTON'S MODEL</u>	8
3. THE GENERALIZED LINEAR MODEL	8
4. OPTIMIZATION STRATEGY FOR THE GENERALIZED LINEAR MODEL	9
4.1 <u>ERRORS IN THE PUBLISHED DATA</u>	11
5. PROBLEMS WITH THE MINIMIZATION, AND THE BEST RESULTS OBTAINED ...	13
6. PREDICTIONS BY VARIOUS MODELS	18
6.1 <u>PREDICTIONS OF THE SANDERS-WYSZECKI BRIGHTNESS DATA</u>	19
6.1.1 Howett, Best-Fitting	19
6.1.2 Howett, Restricted	19
6.1.3 Guth and Lodge, Original	21
6.1.4 Guth and Lodge, Adjusted	24
6.1.5 Thornton, Original	26
6.1.6 Thornton, Adjusted	26
6.2 <u>OPPONENT-LEVEL CHANNEL RESPONSES</u>	29
6.2.1 Guth and Lodge	30
6.2.2 Thornton	30
6.2.3 Howett, Best-Fitting	33
6.2.4 Howett, Restricted	35

TABLE OF CONTENTS (Continued)

	<u>Page</u>
6.3 <u>SATURATION OF THE SPECTRUM</u>	35
6.3.1 Guth and Lodge	37
6.3.2 Thornton	39
6.3.3 Howett, Best-Fitting	39
6.3.4 Howett, Restricted	39
6.4 <u>EQUIVALENT LUMINANCE OF THE SPECTRUM</u>	43
6.4.1 Guth and Lodge	44
6.4.2 Thornton	44
6.4.3 Howett, Best-Fitting	44
6.4.4 Howett, Restricted	44
6.5 <u>WAVELENGTH DISCRIMINATION</u>	49
6.6 <u>NEUTRAL POINTS</u>	54
6.6.1 Guth and Lodge	57
6.6.2 Thornton	59
6.6.3 Howett, Best-Fitting	61
6.6.4 Howett, Restricted	61
6.7 <u>THE PRIMARIES</u>	64
6.7.1 Guth and Lodge	65
6.7.2 Thornton	68
6.7.3 Howett, Best-Fitting	71
6.7.4 Howett, Restricted	75
6.8 <u>CONTOURS OF CONSTANT B/A</u>	75
6.8.1 The Special Case $p=2$ (Including the Guth Model)	78
6.8.2 The Special Case $p=1$ (Including the Thornton Model) ..	79
6.8.3 The General Case (Including the Howett Model)	80
6.8.4 Guth and Lodge	83
6.8.5 Thornton	83
6.8.6 Howett, Best-Fitting	86
6.8.7 Howett, Restricted	89
6.9 <u>ADDITIVITY</u>	89
6.9.1 The Significance of Nonadditivity	89
6.9.2 Vectorial Considerations	92
6.9.3 Definition of a Generalized Nonadditivity Index	92
6.9.4 The Relationship of the Power, p , to Nonadditivity ..	94
6.9.5 Actual Predictions: Pairs of Monochromatic Lights ..	96
6.9.6 Actual Predictions: Lines of Constant Dominant Wavelength	96

TABLE OF CONTENTS (Continued)

	<u>Page</u>
7. SUMMARY AND CONCLUSIONS	102
7.1 <u>SUGGESTIONS FOR FURTHER WORK</u>	105
7.1.1 Some Proposed Experimental Studies	105
7.1.1.1 <u>Is Chromatic Brightness Visually Effective?.</u>	105
7.1.1.2 <u>How Much Coloration in a Light Source Is</u>	
<u>Tolerable?</u>	106
7.1.1.3 <u>The Effect of Viewing Conditions on</u>	
<u>Equivalent Luminance</u>	106
7.1.2 Some Proposed Computational Studies	107
7.1.2.1 <u>Individual Observer Variation</u>	107
7.1.2.2 <u>Studies of Nonlinear Models</u>	107
7.1.2.3 <u>Further Predictions of Color-Vision</u>	
<u>Functions by The Optimized Linear Model</u>	
<u>of This Report</u>	107
8. REFERENCES	108
APPENDIX: BASIC COLORIMETRIC QUANTITIES AND NOTATION	109
<u>CAPITAL LETTERS</u>	109
<u>LOWER-CASE LETTERS</u>	110
<u>LOWER-CASE LETTERS WITH OVERBARS</u>	111

LIST OF TABLES

	<u>Page</u>
Table 1. Null Lines and Primaries of Opponent-Level Channels: Guth and Lodge Model	69
Table 2. Null Lines and Primaries of Opponent-Level Channels: Thornton Model	72
Table 3. Null Lines and Primaries of Opponent-Level Channels: Best-Fitting Howett Model	74
Table 4. Null Lines and Primaries of Opponent-Level Channels: Restricted Howett Model	77

LIST OF FIGURES

	<u>Page</u>
Figure 1. Correlation plot for best-fitting Howett model	20
Figure 2. Correlation plot for restricted Howett model	22
Figure 3. Correlation plot for Guth and Lodge model	23
Figure 4. Correlation plot for adjusted Guth and Lodge model	25
Figure 5. Correlation plot for Thornton model	27
Figure 6. Correlation plot for adjusted Thornton model	28
Figure 7. Opponent-level channel responses for Guth and Lodge model ..	31
Figure 8. Opponent-level channel responses for Thornton model	32
Figure 9. Opponent-level channel responses for best-fitting Howett model	34
Figure 10. Opponent-level channel responses for restricted Howett model	36
Figure 11. Saturation of the spectrum, predicted by Guth and Lodge model	38
Figure 12. Saturation of the spectrum, predicted by Thornton model	40
Figure 13. Saturation of the spectrum, predicted by best-fitting Howett model	41
Figure 14. Saturation of the spectrum, predicted by restricted Howett model	42
Figure 15. Equivalent luminous efficiency function, predicted by Guth and Lodge model	45
Figure 16. Equivalent luminous efficiency function, predicted by Thornton model	46
Figure 17. Equivalent luminous efficiency function, predicted by best-fitting Howett model	47
Figure 18. Equivalent luminous efficiency function, predicted by restricted Howett model	48
Figure 19. Wavelength discrimination function, predicted by Guth and Lodge model	50

LIST OF FIGURES

	<u>Page</u>
Figure 20. Wavelength discrimination function, predicted by Thornton model	51
Figure 21. Wavelength discrimination function, predicted by best-fitting Howett model	52
Figure 22. Wavelength discrimination function, predicted by restricted Howett model	53
Figure 23. Opponent-channel null lines and white point, implied by Guth and Lodge model	58
Figure 24. Opponent-channel null lines and white point, implied by Thornton model	60
Figure 25. Opponent-channel null lines and white point, implied by best-fitting Howett model	62
Figure 26. Opponent-channel null lines and white point, implied by restricted Howett model	63
Figure 27. Opponent-level primaries of Guth and Lodge model	66
Figure 28. Opponent-level primaries of Thornton model	70
Figure 29. Opponent-level primaries of best-fitting Howett model	73
Figure 30. Opponent-level primaries of restricted Howett model	76
Figure 31. Constant-B/A loci, predicted by Guth and Lodge model	84
Figure 32. Constant-B/A loci, predicted by Thornton model	85
Figure 33. Constant-B/A loci, predicted by best-fitting Howett model ..	87
Figure 34. Constant-B/A loci, predicted by restricted Howett model	90
Figure 35. The principle of vector addition	93
Figure 36. Additivity of mixtures of two monochromatic lights, predicted by restricted Howett model	97
Figure 37. Additivity along lines of constant dominant wavelength, predicted by restricted Howett model	100

A NOTE ON NOTATION: UNDERLINING

In standard printing practice, letters representing mathematical variables are printed in italics to distinguish them from ordinary text. Since this report is photoreproduced from a typed manuscript, the customary practice of underlining to represent italics has been followed. To avoid the unnecessary visual clutter of too much underlining, variables that have other letters or symbols attached to them, either vertically or laterally, have, in almost all cases, not been underlined. (This includes variables having subscripts, superscripts, or overbars; or which appear as parts of equations or algebraic expressions.) In short, the intention was to underline only single, unadorned, isolated letters.

Underlining in this report never has any significance other than simply indicating italics. In particular, vectors are represented by capital letters with overbars, and no matrices or tensors are used at all. In no case does an underlined letter differ in meaning from the same letter appearing without the underline, even where the above rules appear to have been violated. (There are almost surely examples in the report of failure to adhere rigidly to these rules.)

All diacritical marks other than underlines are used to distinguish different quantities. Likewise, capital letters always represent variables distinct from those indicated by the corresponding lower case letters. To interpret these distinctions, see the Appendix, which summarizes standard colorimetric notation.

1. INTRODUCTION

It has been known for a long time that if a strongly colored (saturated*) light, such as a red, for example, is examined side by side with a white light that measures the same luminance as the red on a light meter, most people will say that the red appears to them to be considerably brighter. At one time, it was said that the cause of this phenomenon (the Helmholtz-Kohlrausch effect) was that people tended to confuse saturation with brightness.

The predominant current view is that the red looks brighter than the white of equal measured luminance not because of any confusion in people's minds, but because the visual system is organized in such a way that the red really does stimulate the brightness-perceiving center of the brain more than the white does. The class of theory that explains this effect, at least in a qualitative way, is a version of opponent-colors theory. [The essential details of this class of theory are listed in section 1.1. Also, see Howett (in press) for an extended explication of the problem.]

The simplest version of the class of opponent-color theories that explain the Helmholtz-Kohlrausch effect are the linear models. Until now, these linear opponent-color theories have been somewhat successful at explaining additivity problems related to the Helmholtz-Kohlrausch effect, at near-threshold luminance levels, but less successful at modeling these effects at suprathreshold levels. Since most vision in everyday life takes place at luminances well above threshold,

* Saturation is a perceptual variable, indicating, for lights, the perceived degree of difference of the color of the light from an equally bright white light. It can be identified with "strength" or "vividness" of the color. There is no easily specifiable physical correlate of saturation.

photometrists are actively seeking a formula for predicting, with some accuracy, the brightness relations among suprathreshold lights (CIE, 1978).

The present investigation was directed toward deriving such a formula and seeing how well it could predict at least one particular set of experimental results. The basic approach was to write a formula for a generalized version of the linear opponent-colors model, and then to optimize the constants in the formula to obtain the best possible agreement with the behavioral data.

1.1 HISTORICAL BACKGROUND

S.L. Guth has been using a color-vision model of the opponent-colors type to explain why luminance (measured) and brightness (perceived) can disagree strongly when the lights being compared have different chromaticities. The essential features of Guth's model, which he has elaborated in a series of publications for about the last 20 years, are as follows:

- (a) The spectral absorptance curves of the three cone pigments can be expressed as linear combinations of the CIE 1931 color-matching functions, $\bar{x}(\lambda)$, $\bar{y}(\lambda)$, and $\bar{z}(\lambda)$ *. [Most vision researchers accept Judd's (1951) short-wave revision of these functions as closer to correct, but Vos (1978) has shown that Judd's curves can themselves be well approximated by a linear combination of the CIE functions.]
- (b) The spectral response functions of the three opponent-color processes (of which two are opponent and chromatic, and the third is the

* A brief description of the basic quantities of colorimetry and the standard notation is given in the Appendix of this report.

non-opponent luminance channel) arise from reassortment of the outputs of the three cone types in three different combinations. The contribution of each cone output to each of the opponent-colors channels may be either excitatory (positive) or inhibitory (negative).

- (c) It is assumed that no nonlinear processes intervene between the absorption of light within the cone pigments and the generation of output within the opponent-colors channels. It is in this sense that the model is "linear". As a result, the outputs of the opponent-colors channels are themselves linear combinations of the CIE 1931 color-matching functions.
- (d) Guth then assumes that the outputs of the opponent-colors channels are further processed as if they were independent dimensions of a three-dimensional euclidean vector space. In Guth's own terms, his model is a vector model. Any color can then be thought of as a vector from the origin, with its components equal to the outputs of the three opponent-colors channels.
- (e) The length of this color vector, calculated by the usual euclidean rule--square root of the sum of the squares of the components--is called "vector luminance" by Guth, and is identified by him as a measure that does agree with brightness judgments, regardless of chromaticity, at least under conditions at or near threshold detection.

A model very similar in structure to Guth's, differing only in its quantitative details, has been under development by Ingling and his coworkers (Ingling and Tsou, 1977; Ingling, 1977) for a comparable period of time. Since the Ingling

model introduces no new basic features for purposes of the present investigation, only minimal further reference will be made to it. This omission should not be construed as reflecting at all on the relative merits of the Guth and Ingling models.

Later, Thornton (1973) accepted the basic concepts of Guth's model, but changed the opponent channel responses, as well as the final formula for combining the components of the color vector. Thornton reasoned that nerve cells tend to combine their outputs in a direct fashion, in which the total stimulation of a cell is simply the sum of the stimulations received from the input channels. He therefore took the metric of his vector space to be simply the sum of the absolute values of the components (the opponent-colors channel outputs), in contrast to the square root of the sum of squares used by Guth and by Ingling.

2. THE NUMERICAL MODELS

2.1 GUTH'S MODEL

Various versions of the Guth model have been published by Guth and an assortment of coworkers over the years. In a recent version (Guth, Massof, and Benzscharow, 1980), a great deal of flexibility in the form of six adjustable constants was introduced. These constants represent allowances for the process of chromatic adaptation as applied separately to the cone responses and to the opponent-level channel responses. Although adjustments for adaptation must undoubtedly be part of a complete color-vision model, the Guth-Massof-Benzscharow treatment does not explicitly express these parameters as functions of luminance, chromaticity, or other measurable variables; they are in practice arbitrary constants to be fitted to data by trial and error. Since there was no obvious basis for selecting any particular values for the adjustable constants, the

version that follows was chosen from an earlier publication (Guth and Lodge, 1973), in which the model still had a fixed format.

For purposes of this analysis, there was no need to separate steps (a), (b), and (c) of section 1. Mathematically, applying a linear transformation to another linear transformation yields still another linear transformation. Therefore, the step referring to the cone responses is omitted, and the formulas for the opponent-colors channels are the starting point of the treatment. These formulas, for the Guth and Lodge (1973) model, are:

$$\begin{aligned}\bar{a} &= 0.954\bar{y} + 0.010\bar{z} \\ \bar{t} &= 0.799\bar{x} - 0.646\bar{y} - 0.167\bar{z} \\ \bar{d} &= -0.058\bar{y} + 0.030\bar{z}.\end{aligned}\tag{1a}$$

In Eqs. (1a), $\bar{x}$, $\bar{y}$, and $\bar{z}$ are the CIE 1931 color-matching functions, and are functions of wavelength, as are the opponent-colors responses $\bar{a}$, $\bar{t}$, and $\bar{d}$ (Guth's notation). Thus, $\bar{x}$, for example, would be more explicitly written as $\bar{x}(\lambda)$, but the conventional use of overbars to denote color-matching functions (formerly also called spectral distribution coefficients) makes it unambiguous to omit the parenthetical λ in most contexts.

In applying the model to specific colors, $\bar{x}$, $\bar{y}$, and $\bar{z}$ are replaced by the general tristimulus values $\underline{X}$, $\underline{Y}$, and $\underline{Z}$, on the right sides of Eqs. (1a), and $\bar{a}$, $\bar{t}$, and $\bar{d}$ are replaced by tristimulus values $\underline{A}$, $\underline{T}$, and $\underline{D}$ on the left sides.

The resulting equations are

$$\begin{aligned}A &= 0.954Y + 0.010Z \\ T &= 0.799X - 0.646Y - 0.167Z \\ D &= -0.058Y + 0.030Z\end{aligned}\tag{1}$$

$$B = (A^2 + T^2 + D^2)^{1/2}.\tag{2}$$

In Eqs. (1), $\underline{A}$, $\underline{T}$, and $\underline{D}$ are no longer functions of wavelength but apply to the single color of which $\underline{X}$, $\underline{Y}$, and $\underline{Z}$ are the tristimulus values. It is not a coincidence that the coefficients in Eqs. (1) and (1a) are identical. The quantities $\bar{x}$, $\bar{y}$, and $\bar{z}$ of Eqs. (1a) are themselves tristimulus values, of the monochromatic lights comprising the visible spectrum (with equal energy at all wavelengths). In fact, the currently approved name (CIE, 1970) for these quantities is "tristimulus values of the spectrum." The same quantities are also properly referred to as "color-matching functions" because [see the Appendix], in any colorimetric system, the tristimulus values of the spectrum determine what light spectra match each other visually.

Since Eqs. (1) are fully general, they apply to any color; in particular, they apply to monochromatic lights. Thus, Eqs. (1a) apply to a subset of the stimuli to which Eqs. (1) apply, and hence the two sets of equations must have the same coefficients. This is a general rule, applying to transformations between any two colorimetric systems that have a linear relationship to each other. As a result of the validity of this rule, the two sets of equations, (1) and (1a), are thought of as equivalent and are often used interchangeably in publications. It is then left to the context to communicate to the reader whether equations having the form of (1) are intended to refer to spectral functions or the tristimulus values of individual colors.

In Eqs. (1a), $\bar{a}$, $\bar{t}$, and $\bar{d}$ represent the spectral outputs of the opponent-colors channels. The $\underline{A}$ channel [the more common notation based on Eqs. (1)] is non-opponent and is the luminance, whiteness, or achromatic channel. The $\underline{T}$ channel is usually described as the red-green opponent channel, in which red colors (high $\underline{X}$) yield positive stimulation ($T > 0$), and green colors (high $\underline{Y}$) yield

negative stimulation ($T < 0$). Similarly, the D channel is usually described as the blue-yellow channel, in which blue colors (high Z) yield (in Guth's treatment) positive stimulation ($D > 0$), and yellow colors (high X and Y) yield negative stimulation ($D < 0$). The choice of sign in constructing these channels is arbitrary, and other authors prefer to make yellow the positive phase of the D channel.

The quantity B in Eq. (2) is, in Guth's terms, the vector luminance because it is derived from his vector model; B is itself a scalar quantity. B is not Guth's notation, but that letter is commonly used in other publications on this subject. It is often referred to as "brightness," but is more accurately described as a measure that correlates well with brightness. Brightness is itself a subjective sensation level (CIE, 1970). Operationally, the B value of a light can be identified as the luminance of a reference white light that matches the light in question in brightness. The name used for B in this report is "equivalent luminance" (CIE, 1970), following the convention introduced by the author in an earlier publication (Howett, in press).

In the current system of photometry, luminance is determined entirely by the value of Y, or, in the Guth model, the near-equivalent A. The essence of the explanatory power of Guth's model is that the quantity B includes not only the luminance A, but also contributions from the chromatic (and opponent) channels T and D. The result is that strong (saturated, vivid) colors which have significant values of T and/or D, have their equivalent luminances (B) enhanced by these contributions from the chromatic channels. This extra contribution is often referred to as "chromatic brightness," and is a feature of virtually all contemporary opponent-colors models, linear or nonlinear.

2.2 THORNTON'S MODEL

The following model is due to Thornton (1973). It is presented using notation consistent with Eqs. (1) and (2), rather than in Thornton's notation. The tristimulus-value version is given.

$$\begin{aligned} A &= Y \\ T &= 0.4X - 0.4329Y + 0.2073Z \\ D &= 0.4X - 0.4322Y - 0.2229Z \end{aligned} \tag{3}$$

$$B = |A| + |T| + |D|. \tag{4}$$

The nonopponent A channel is always positive, so that the absolute value sign around A in Eq. (4) is not really necessary.

Aside from the difference in the rule for computing B, Thornton has definitions of his opponent channels (T and D) quite different from those used by Guth. The derivation of these definitions arose from some different considerations in Thornton's treatment than in Guth's analysis, but here we are concerned only with the mathematical form, and not, at this stage, with the theoretical interpretation.

3. THE GENERALIZED LINEAR MODEL

There are obvious similarities between the structures of the models embodied in Eqs. (1)-(2) and Eqs. (3)-(4). First, Eqs. (1) and (3) both have the form

$$\begin{aligned} A &= a_{11}X + a_{12}Y + a_{13}Z \\ T &= a_{21}X + a_{22}Y + a_{23}Z \\ D &= a_{31}X + a_{32}Y + a_{33}Z, \end{aligned} \tag{5}$$

where the a_{ij} ($i, j=1, 2, 3$) represent nine general constants.

Slightly less obvious is that the composition rules for B given by Eqs. (2) and (4) are also of the same basic form. Both are special cases of the equation

$$B = (|A|^p + |T|^p + |D|^p)^{1/p}, \quad (6)$$

where p is a general constant. The metric represented by Eq. (6) is well known in mathematics, and is often referred to as the Minkowski L_p metric. For Guth [Eq. (2)], $p = 2$; and for Thornton [Eq. (4)], $p = 1$.

It is necessary to take absolute values of the channel outputs in Eq. (6) because fractional values of p are possible, and these cannot in general be applied as exponents to negative numbers. There was no need for the absolute value signs in Eq. (2), since squaring a number performs the equivalent function of making the sign of the number irrelevant.

4. OPTIMIZATION STRATEGY FOR THE GENERALIZED LINEAR MODEL

The model represented by Eqs. (5) and (6) contains 10 arbitrary constants (parameters): the nine a_{ij} values of Eq. (5), and p of Eq. (6). The goal of this study was to select a body of heterochromatic brightness-matching data and optimize the 10 constants of the model, in the sense of making the predictions of the model concerning equivalent luminances agree best with the data.

Two strategies were available in selecting a data set to work with. One was to average data from several different studies; the other, to choose one extensive, trustworthy, internally consistent body of data. The trouble with the various significant heterochromatic brightness-matching studies in the literature --perhaps 8 to 10 in number-- is that they were all carried out with different viewing conditions (e.g., area, luminance, separation, presentation time). As a result, although the general pattern of the results is similar in all the

reliable studies, the sizes of the numbers are very different. The author's opinion is that averaging such disparate data tends to obscure rather than reveal the underlying structure of the situation. Therefore, the option chosen was the single-experiment strategy, with the hope of successfully predicting the data for one particular set of viewing conditions. An experimental study that it seems important to carry out in the future is the systematic exploration of the effects of varying these viewing parameters separately and in combination.

The experiment chosen was one carried out at the National Research Council of Canada (Sanders and Wyszecki, 1964). Ninety-six colors scattered over the chromaticity diagram, all having a fixed luminance of 20 cd/m^2 , were matched in brightness by adjustment of a reference white light, by 20 observers. The luminance of the matching white light --i.e., the equivalent luminance of the colored light-- was the dependent variable of the experiment. For each of the colored lights, the experimental data can be embodied in three numbers: the physically measured $\underline{X}$ and $\underline{Z}$ tristimulus values ($\underline{Y}$ being constant at 20), and the psychophysically (experimentally) determined equivalent luminance, B^* . What was optimized, in deriving the best-fitting model, was the agreement between the empirically determined values, B^* , and the model predictions, $\underline{B}$, of Eq. (6). Specifically, the goal was to minimize the sum of the squared prediction errors,

$$S = \sum_{i=1}^{96} (B_i - B_i^*)^2. \quad (7)$$

The plan was to find those values of the a_{ij} in Eq. (5) and the value of $\underline{p}$ in Eq. (6) that minimized the objective function, $\underline{S}$, of Eq. (7).

4.1 ERRORS IN THE PUBLISHED DATA

Several errors in the published tables of the Sanders and Wyszecki (1964) paper have been uncovered. Three of these are typographical errors in the y chromaticity coordinates of the samples. Since the tables contain CIE 1960 u,v as well as x,y values, there was no difficulty in deriving the correct y values, on the assumption that there were not also errors in u or v. Because all the colors used had chromaticities lying along a series of straight, radial lines in the chromaticity diagram, it was also possible to closely estimate the value of y by linear interpolation between the coordinate of the two points on either side on the same line, using the x values to determine the proportion. The values of y derived from u, v and from linear interpolation agreed well in all three cases.

The following table shows the correct and incorrect values:

<u>Dominant Wavelength</u> <u>/Purity Level</u>	<u>Published</u> <u>y Value</u>	<u>Correct</u> <u>y Value</u>
583/4	0.134	0.413
574/3	0.422	0.442
450/4	0.277	0.227

It can be seen that in all three cases, the nature of the error is the transposition of digits, or the repetition of the wrong digit. It is easy to visualize how such errors can be made in the process of setting type from the original material supplied by the authors. The other group of errors (found by Robert Clear, Lawrence Berkeley Laboratory), is more difficult to explain. The published mean logarithm of what Sanders and Wyszecki refer to as B/Y values (equivalent luminance divided by luminance) is in two cases not actually equal to the mean of the 20 logarithms for the individual subjects. The following table shows these discrepancies.

<u>Dominant Wavelength /Purity Level</u>	<u>Published Mean Log B/Y Value</u>	<u>Correct Mean Log B/Y Value</u>
583/5	0.058	0.068
487/1	0.018	0.012

In both of these cases, the error is in a single digit and could also be the result of careless typesetting, but the source of these errors is somewhat uncertain. One defensible reaction to this problem would be simply to omit these two colors from the analysis. This is the conservative strategy that is above reproach. On the other hand, these difficult-to-obtain data should not be discarded lightly. In order for the mean logs to be off by as much as they are, it is probable that errors in more than one of the individual values would have had to be made, since no one individual value is obviously grossly deviant. Moreover, in contrast with the printed table of mean values, the published table of the individual log values is clearly a photocopy of a typed table that presumably was calculated and checked by the authors of that paper before it was submitted. Consequently, the chances are good that the error was made in the typesetting of the mean values, and not in the specification of the individual values, so that the choice of adopting the actual means of the individual values as correct is also defensible, from a probabilistic standpoint.

Actually, it most likely makes little difference whether a few points are omitted from the analysis, or even whether a few slightly (but not grossly) erroneous points are included. This is because the spread (variance) of the data is considerable, and because--as will be described later--the optimization of the fitted model is so difficult.

5. PROBLEMS WITH THE MINIMIZATION, AND THE BEST RESULTS OBTAINED

Although the optimization problem posed by the set of equations (5)-(7) appears deceptively simple in form, it has proved to be extraordinarily difficult in practice. The first attempt was made using an optimization routine included in a commercial package of computer subroutines, usually regarded as of high overall quality. Ideally, in applying an optimization procedure, one hopes that the same optimum point will always emerge, regardless of the starting point, the initial step size, and the other parameter values that can be varied. Such behavior instills confidence that the true global minimum has been found.

It quickly became apparent that the situation with this minimization problem was very far from this ideal. In fact, a new alleged minimum point (in 10-dimensional space) emerged with virtually every different choice of starting parameters. It appears that the function $\underline{S}$ of Eq. (7) has a tremendous number of local minima. As the starting point and initial step size were varied, values of $\underline{S}$ lower than any previously discovered would turn up occasionally. It became clear that it would be difficult to find the overall, global minimum, and that even if, by chance, the global minimum were located, it would be difficult to conclude that it was in fact the true global minimum.

Since the package routine was not succeeding, the present author decided to program an optimization algorithm of his own that was of a conceptually simple, brute-force type. It seemed, in advance, that the algorithm might require many iterations, but that it might succeed in getting to the global minimum where the more sophisticated package routine had failed. Unfortunately, the result of applying the new program was very similar to the outcome of the attempt using

the package routine: a sequence of many different minima, with a new, lower minimum emerging from time to time through trial and error.

A team of two mathematicians with special skills in optimization theory was consulted regarding the difficult behavior of this particular function. They were unable to find a global minimum even with the most advanced algorithms available to them, and indeed did not locate a local minimum as good as the best one the author had found in his repeated attempts. Apparently, this particular problem presents some sort of exceptional and unexpected challenge to existing optimization theory and techniques. One factor that complicates matters is the use of absolute values, which leads to sharp "points" or "spikes," rather than the more tractable smooth "domes" (in 10 dimensions). Another possible contributing factor is that there may be interdependencies among the 10 variables, so that the number of independent parameters in the true underlying problem may be less than 10; i.e., the present formulation of the problem may be overdetermined.

Although for the above reasons a definitively best set of values cannot be presented here, the best local optimum that the author has found predicts the brightness data very well. It might really be the global optimum, or possibly a slightly lower minimum will some day be found. In any event, the formula presented below does establish two key points, which may be regarded as the principal contributions of the present study: (a) it is possible to predict a single body of heterochromatic brightness-matching data quite well with a "linear" opponent-colors model; and (b) the use of formal, multivariate optimization procedures can --in combination with some trial and error, as

required-- produce useful and even surprising results, as opposed to the intuitive, pure trial-and-error procedures that have often been used in the past.

The best-fitting prediction of the Sanders-Wyszecki (1964) data that was found is given by the following formula:

$$\begin{aligned} A &= -0.0647X + 1.0583Y - 0.1294Z \\ T &= 0.2914X - 0.2341Y - 0.0312Z \\ D &= -0.1028X + 0.1572Y - 0.1639Z \end{aligned} \quad (8)$$

$$p = 0.8184. \quad (8a)$$

The quantity $\underline{p}$ of Eq. (8a) is used to predict equivalent luminance, $\underline{B}$, in accordance with Eq. (6). The sum of the squared errors of prediction, $\underline{S}$, from Eq. (7), is only 36.047 for the 96 colors. The mean absolute prediction error for a single color is 0.480, or 1.938 percent of the overall mean $\underline{B}^*$ value. The product-moment correlation, $\underline{r}$, is 0.986.

It is important to note that the formula embodied in Eqs. (8) and (8a) was obtained without the imposition of any restrictions whatever. The sole criterion was minimizing the sum of the squared errors of prediction of the experimentally determined equivalent luminances. However, despite the determination of the formula exclusively by brightness data, Eqs. (8) represent a set of opponent-color functions that resemble considerably the functions presented by other authors. This promising finding encouraged the hope that a slight modification of Eqs. (8)-(8a) might permit reasonable predictions of some of the classical color-vision functions, such as the predictions exhibited by Guth and co-workers.

The most obvious defect of Eqs. (8) --as will be shown in a later section-- is that the luminance function $\underline{A}$ is negative in the shortwave (blue) extreme of the spectrum. A valid luminance function cannot be negative anywhere, since pure light of any wavelength (or mixture of wavelengths) at which the function

is negative would have to have negative luminosity, a possibility that experience establishes as unrealistic. A second optimization was therefore carried out subject to the single restriction that $\underline{A}$ be nowhere negative. In practice, this restriction was simplified computationally to the triple limitation that all three of the coefficients in the linear expression for $\bar{a}$ as a function of $\bar{x}$, $\bar{y}$, and $\bar{z}$ must be non-negative. This guarantees an all-positive $\bar{a}$ function, but the condition is excessively restrictive; i.e., all-positive $\bar{a}$ functions can exist in principle even when the coefficient of $\bar{x}$ or $\bar{z}$ is slightly negative. The restriction to a nowhere-negative luminance function is, at least on the surface, still confined to brightness-related considerations. Nevertheless, as will be seen later, the model that emerges is not too bad at predicting color-related functions.

The best model found under the restriction described above was:

$$\begin{aligned} A &= 0.0015X + 0.9865Y \\ T &= 0.2691X - 0.2056Y - 0.0431Z \\ D &= -0.0606X + 0.0858Y - 0.0582Z \end{aligned} \tag{9}$$

$$p = 0.8174. \tag{9a}$$

The model embodied in Eqs. (9) and (9a) is not extremely different from the totally unrestricted model incorporated in Eqs. (8) and (8a). Despite the restriction imposed on $\bar{a}$, this model predicts the brightness-matching data nearly as well as the unrestricted model does. The sum of the squared prediction errors is a little higher, at 43.635; but the mean absolute error is actually less, at 0.470 (1.896 percent). The correlation is 0.983, negligibly lower.

It should be pointed out that the coefficient of $\underline{X}$ in the equation for $\underline{A}$ given in Eqs. (9) is quite small. During the computer run that led to the particular

set of (at least locally) optimum values shown in Eqs. (9) and (9a), the coefficient of $\underline{Z}$ in the expression for $\underline{A}$ stayed equal to zero (meaning that it was trying to go negative), but the coefficient of $\underline{X}$ kept a slightly positive value. Nevertheless, there would probably be little consequence numerically if the $\underline{X}$ term were simply dropped and $\underline{A}$ set equal to $\underline{Y}$. A desirable step, which has not been carried out, would be to make $\underline{A}$ proportional only to $\underline{Y}$, (i.e., to choose $a_{11} = a_{13} = 0$ in advance), and then re-optimize the resulting 8-variable function.

The calculations of the predictions of this model, to be described later, were based on the values as given in Eqs. (9). In a few cases, the predictions were recomputed with the $\underline{X}$ term of the equation for $\underline{A}$ (i.e., a_{11}) set equal to zero, and the differences in the predicted curves were not noticeable, but not all of the predicted functions were checked in this way. There is a statistical test which can be applied to determine whether the predictions of equivalent luminance made with Eqs. (9) are or are not significantly better than the predictions made by setting $a_{11} = 0$. Such tests were not carried out, however, because:

(a) the statistics are based on theoretical considerations properly applicable to linear least-squares problems, and the degree of validity of these tests as applied to nonlinear least-squares (as this problem would be classified, even though the model is labeled "linear") is not well understood; and (b) there is so much "fuzziness" in the data and in the optimization process that expending computational effort on statistical tests did not seem justified, in comparison with using the time to explore another predicted color function. In the future, it would be of interest to explore the use of statistical tests for deciding between models that fit data almost equally well.

The most important result of the two optimizations [Eqs. (8)-(8a) and (9)-(9a)] is that accurately predicting brightness data requires a power, p , considerably less than unity; in both cases, p is about 0.82. Thus, with respect to the power in the combination rule, Thornton ($p=1$) was closer to the truth than Guth ($p=2$), but even Thornton did not go far enough down to obtain a good fit.

6. PREDICTIONS BY VARIOUS MODELS

This section is devoted to showing, through a series of graphs, the predictions by a group of models with respect both to brightness and to color-perception functions. The prediction of color functions involves additional assumptions beyond the specification of the opponent-level channel responses. It is necessary to postulate, for each perceptual function to be predicted, a specific mathematical rule for combining the opponent-level responses so as to match the desired response curve reasonably well. Rather than introducing new combination rules tailored to make the predictions of the author's restricted model [Eqs. (9)-(9a)] look as good as possible, the decision was made in advance to accept without modification Guth's formulations of how to predict the various functions. This procedure gave the advantage to the Guth model in the competition among the models compared, with respect to the prediction of color functions. It was felt that if the author's model could make a creditable showing under this sort of restriction, there would be hope for its potential for providing a general color-vision model, with possible further adjustment. In the present paper, no modifications of the model are explored, but some comments are made regarding new channel-combination formulas that can be examined in the future for possibly improved predictions of some of the color functions.

6.1 PREDICTION OF THE SANDERS-WYSZECKI BRIGHTNESS DATA

This section presents correlation (scatter) diagrams showing predictions by various mathematical models (ordinate) against the experimentally obtained values for the Sanders-Wyszecki (1964) study (abscissa). It should be recalled that the experimental values ["equivalent luminances," in the author's terminology (Howett, in press)] are luminances of a white light seen as having a brightness equal to that of each of the 96 colors tested, all of the latter having been presented at a fixed luminance of 20 cd/m^2 ($\approx 5.8 \text{ fL}$). Thus, the vertical placements of the points in the scatter plots are a function of the particular model, whereas the horizontal placements are fixed, in accordance with the experimental results. If current photometry were correct --i.e., if luminance were perfectly correlated with brightness-- then all the points in each diagram would lie along a vertical line at an abscissa value of 20. Instead, in a now well-established pattern, 20 is the lower bound of the equivalent luminances, and, for this particular experiment, the values range up to approximately 37.

6.1.1. Howett, Best-Fitting

Figure 1 shows the author's best-fitting model, as embodied in Eqs. (8)-(8a). As previously mentioned, the model predicts brightness very well (product-moment correlation $r=0.986$), but involves an unrealistic feature: a luminance function that contains negative values for a range of short wavelengths. The latter feature will be exhibited later, in figure 9.

6.1.2 Howett, Restricted

The re-optimization [see Eqs. (9)-(9a)] was restricted to having an all-positive luminance function. The prediction of brightness in the Sanders-Wyszecki

MODEL: HOWETT, BEST-FITTING

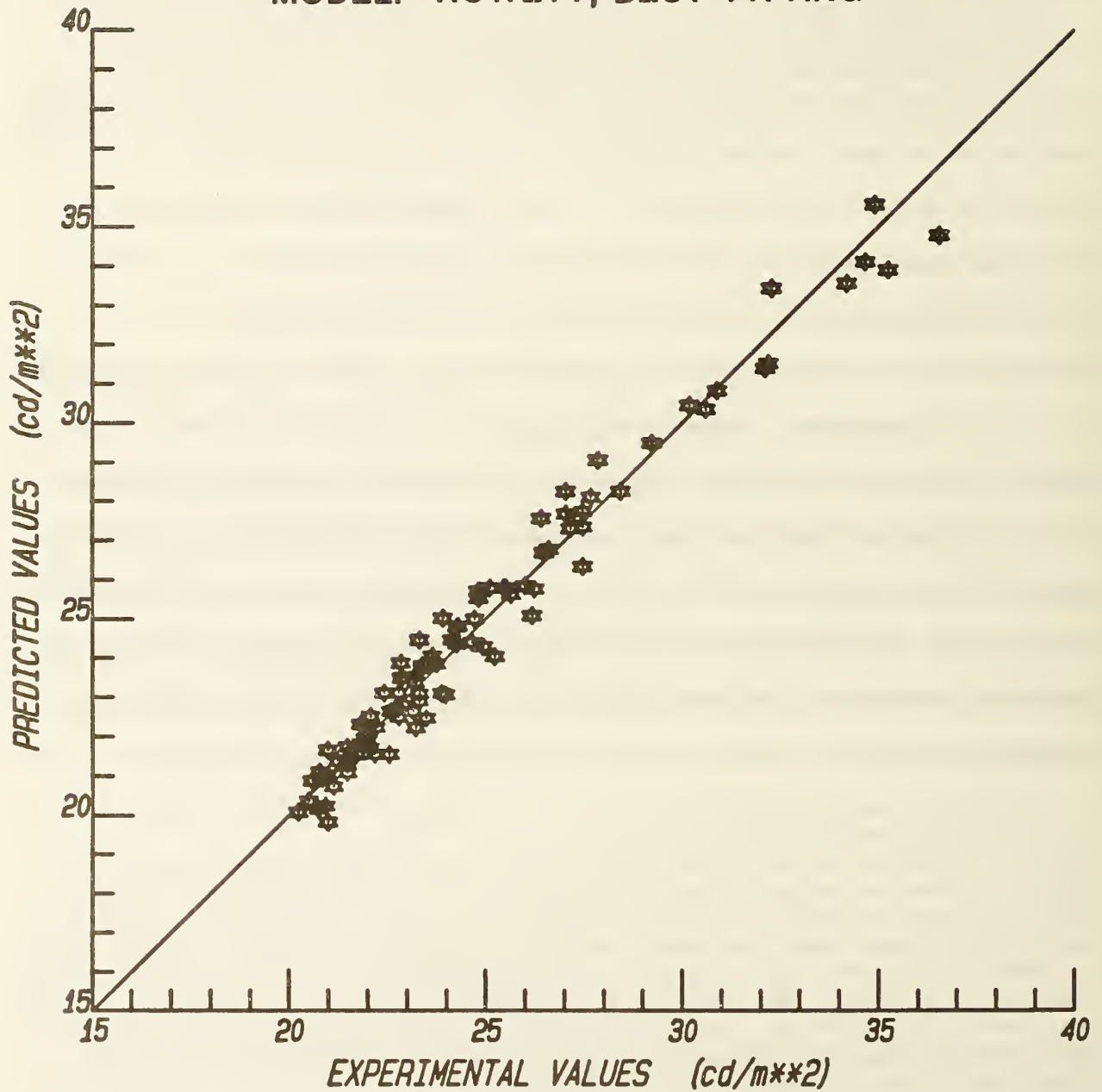


Figure 1. Correlation plot (scatter diagram) for best-fitting Howett model. Predicted equivalent luminance is plotted against the equivalent luminance measured in the Sanders-Wyszecki (1964) experiment.

experiment is still very accurate ($r=0.983$), as seen in figure 2, but the model makes better sense physiologically. Although some predictions from the author's "Best-Fitting" model will be exhibited in later sections of this paper, it is this ("Restricted") model that should be regarded as having some potential for development into a general model of color vision.

6.1.3 Guth and Lodge, Original

Figure 3 shows the predictions of the last version of Guth's model (Guth and Lodge, 1973) that did not contain adjustable constants. Guth's evolving model, in its various forms, has been the most widely accepted model within the vision/color academic community. It was singled out for discussion in CIE Publication No. 41 (1978).

As seen in figure 3, this model fails to predict the Sanders-Wyszecki data well in three ways: (a) the predicted luminances are uniformly too low; (b) for colors having low experimental equivalent luminances (in the 20-24 cd/m^2 range), the model predicts a virtually constant luminance just under 20 cd/m^2 ; and (c) the relationship between predicted and experimental values is highly curvilinear.

It was to be expected that this model would not fit the Sanders-Wyszecki data as well as the author's models, since the brightness data that the Guth model was intended to fit was from a different experiment, conducted near threshold levels. However, the gross discrepancy between data and model seen in figure 3 may seem rather surprising. Guth and his co-workers became aware some time ago that their fixed model did not do well at predicting supra-threshold brightness data, so that they proceeded to develop a generalized model with adjustable

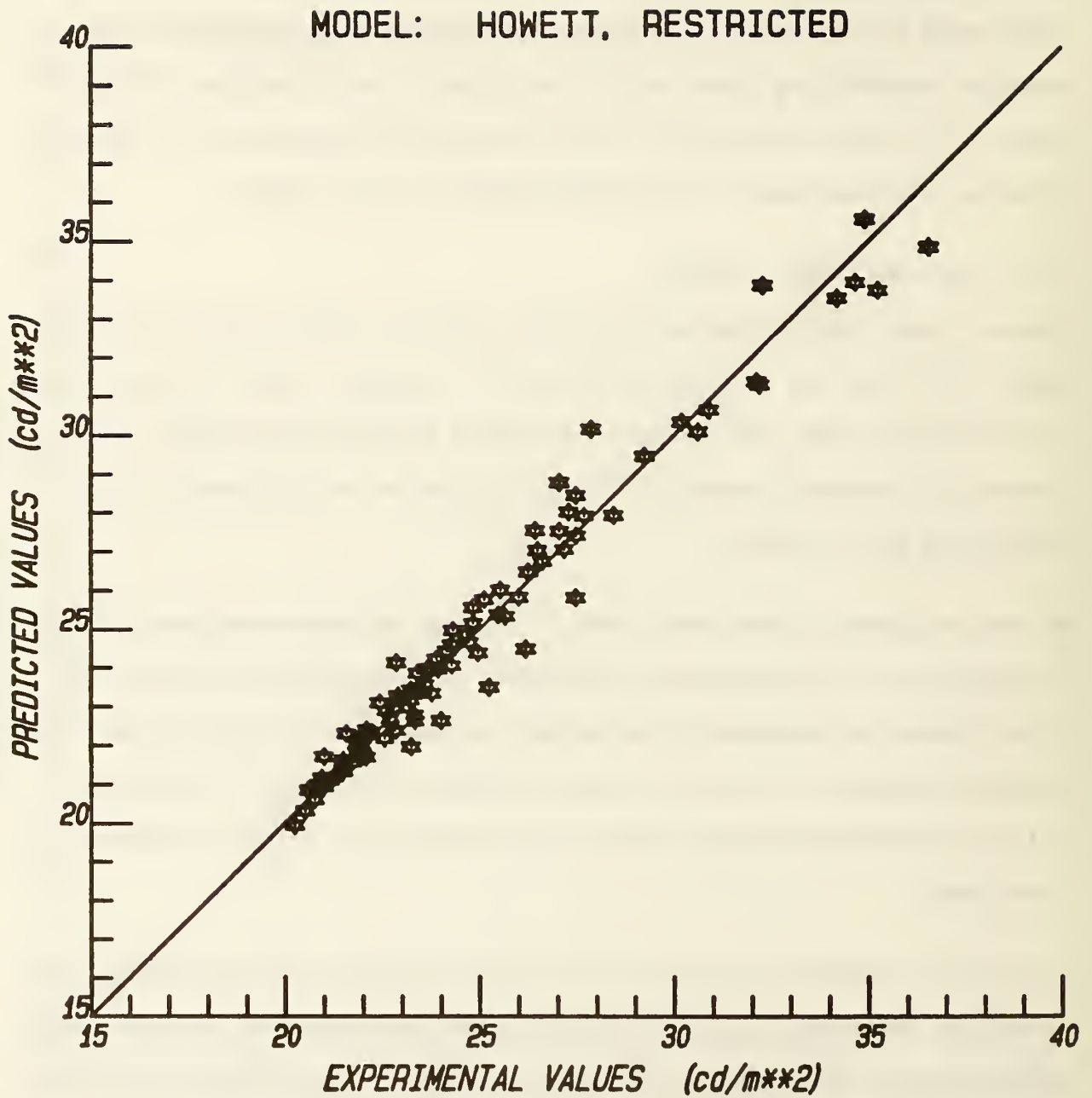


Figure 2. Correlation plot (scatter diagram) for restricted Howett model. Predicted equivalent luminance is plotted against the equivalent luminance measured in the Sanders-Wyszecki (1964) experiment.

MODEL: GUTH & LODGE, ORIGINAL

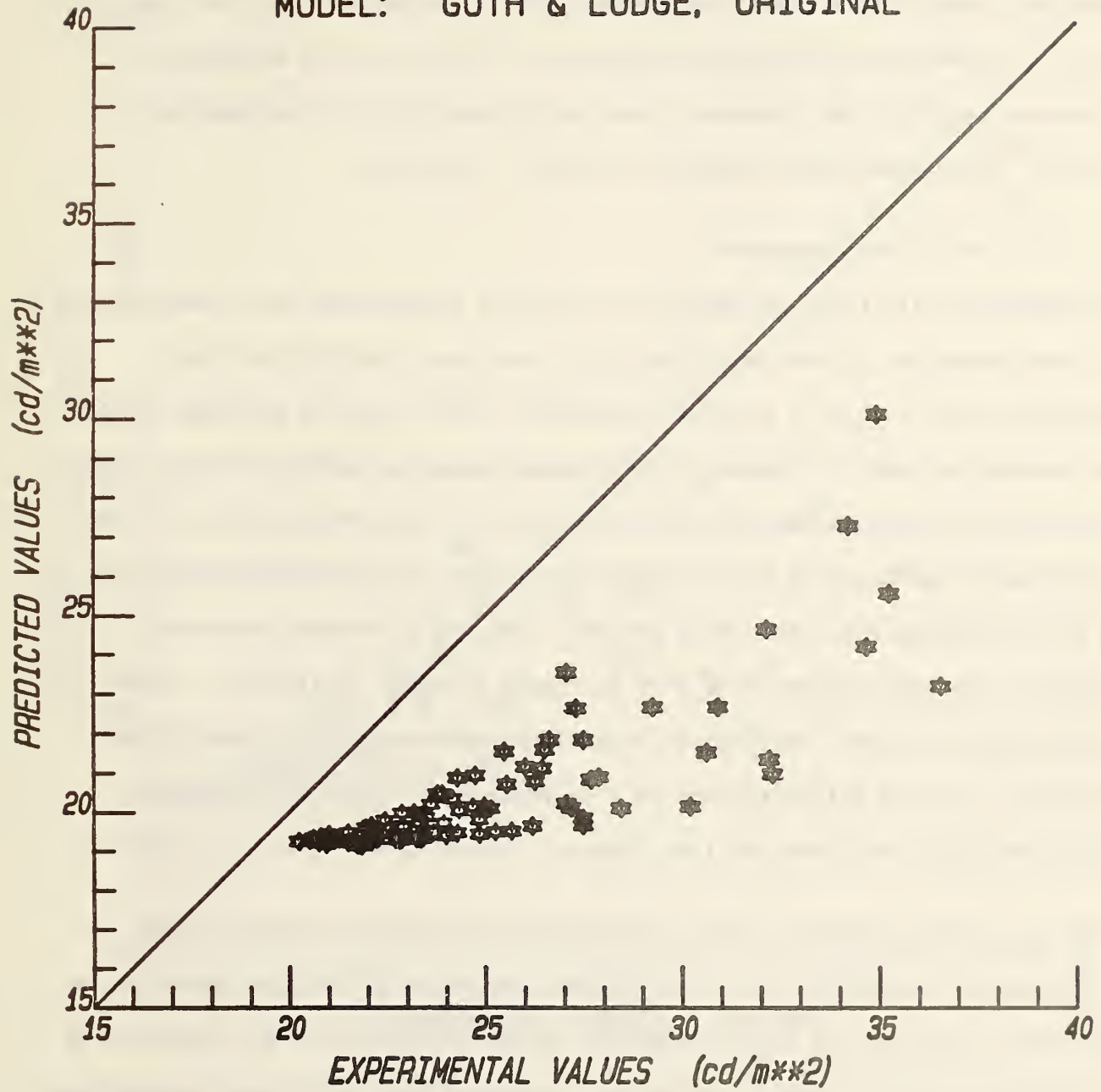


Figure 3. Correlation plot (scatter diagram) for original Guth and Lodge model. Predicted equivalent luminance is plotted against the equivalent luminance measured in the Sanders-Wyszecki (1964) experiment.

parameters to allow for different conditions of luminance and chromatic adaptation (Guth, Massof, and Benzscharwel, 1980). Unfortunately, they have to date provided only qualitative guidance as to how their six adjustable parameters vary with the luminance level and chromaticity of the adapting light, or other objectively measurable stimulus conditions.

6.1.4 Guth and Lodge, Adjusted

The difficulty (a) listed in section 6.1.3 can be compensated for by multiplying all three equations in the model [Eqs. (1)] --or equivalently, the final predictions [Eq. (2)]-- by a suitable constant. The result of carrying through this process is shown in figure 4, the constant selected having the value that gives the least-squares best fit to the data. Only the vertical scale of the predictions is affected by this constant adjustment, and the difficulties (b) and (c) of section 6.1.3 are still present. The sum of squared errors of prediction has been reduced from 2455 in figure 3 to 511 in figure 4. Since linear (product-moment) correlation is unaffected by applying a linear transformation --such as multiplication by a constant-- to either set of values being correlated, the value of $\underline{r}$ in figure 4 is 0.816, the same as in figure 3.

It is interesting that the value of the constant factor that optimizes the sum of the squared prediction errors is not what one might intuitively guess --i.e., the ratio of the mean of the experimental values to the mean value predicted by the unadjusted model-- but is defined by a somewhat more complicated expression, namely

$$k = \frac{\sum_{i=1}^{96} B_i B_i^*}{\sum_{i=1}^{96} B_i^2}, \quad (10)$$

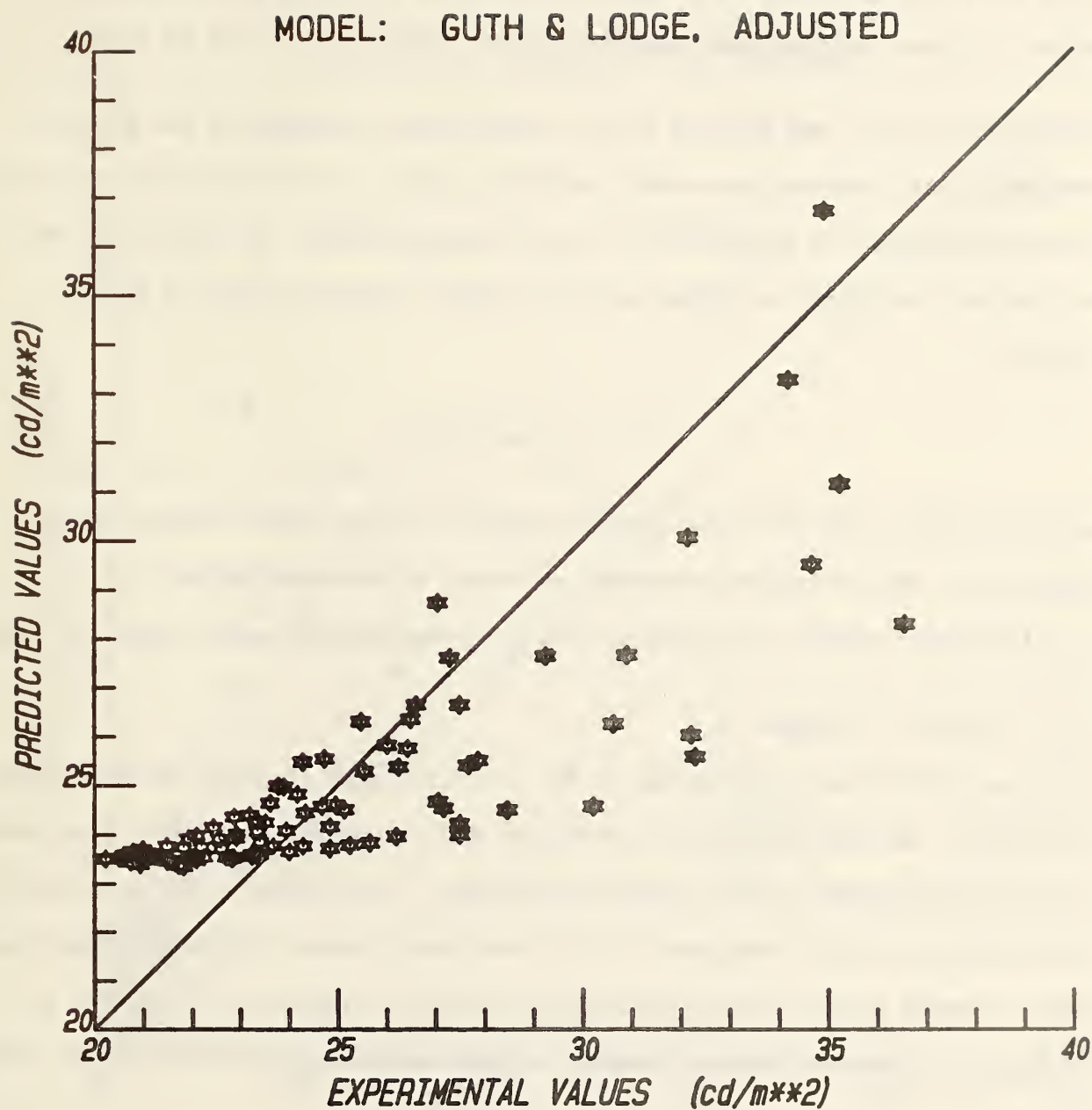


Figure 4. Correlation plot (scatter diagram) for adjusted Guth and Lodge model. Predicted equivalent luminance is plotted against the equivalent luminance measured in the Sanders-Wyszecki (1964) experiment. The absolute scale of the Guth and Lodge predictions (Fig. 3) was adjusted for optimum fit.

where B_i is the i th original predicted value, and B_i^* is the i th experimental value. In this instance, the numerical value of k turned out to be 1.22.

The expression (10) was derived as the least-squares solution to the general problem of what constant one should multiply a given set of values by, in order to best approximate a second set of values. Specifically, if $\{B_i\}$ is any set of n values, and $\{B_i^*\}$ any other set of n values, then the value of k that minimizes

$$U = \sum_{i=1}^n (kB_i - B_i^*)^2 \quad (10a)$$

is given by Eq. (10) [with the specific value 96 of Eq. (10) replaced by the general n]. The derivation proceeds, as usual, by differentiating U of Eq. (10a) with respect to k , and setting the resulting derivative equal to zero.

6.1.5 Thornton, Original

Thornton's (1973) model, embodied in Eqs. (3)-(4), also fits the Sanders-Wyszecki data badly. As seen in figure 5, virtually all the predictions are overestimates, and the relationship is also highly curvilinear. One gathers from his publications that Thornton intended his model to fit "everyday" (supra-threshold) brightness data; it would appear that he derived his formula to agree with some sort of average of a number of heterochromatic brightness-matching studies that he cites.

6.1.6 Thornton, Adjusted

The overall scale of Thornton's model was adjusted by the best-fitting constant, k , derived in accordance with Eq. (10), but figure 6 shows that the predictions are still bad. The sum of the squared errors of prediction was reduced from 5749 in figure 5 to 1925 in figure 6 by multiplying the predictions by $k=0.796$.

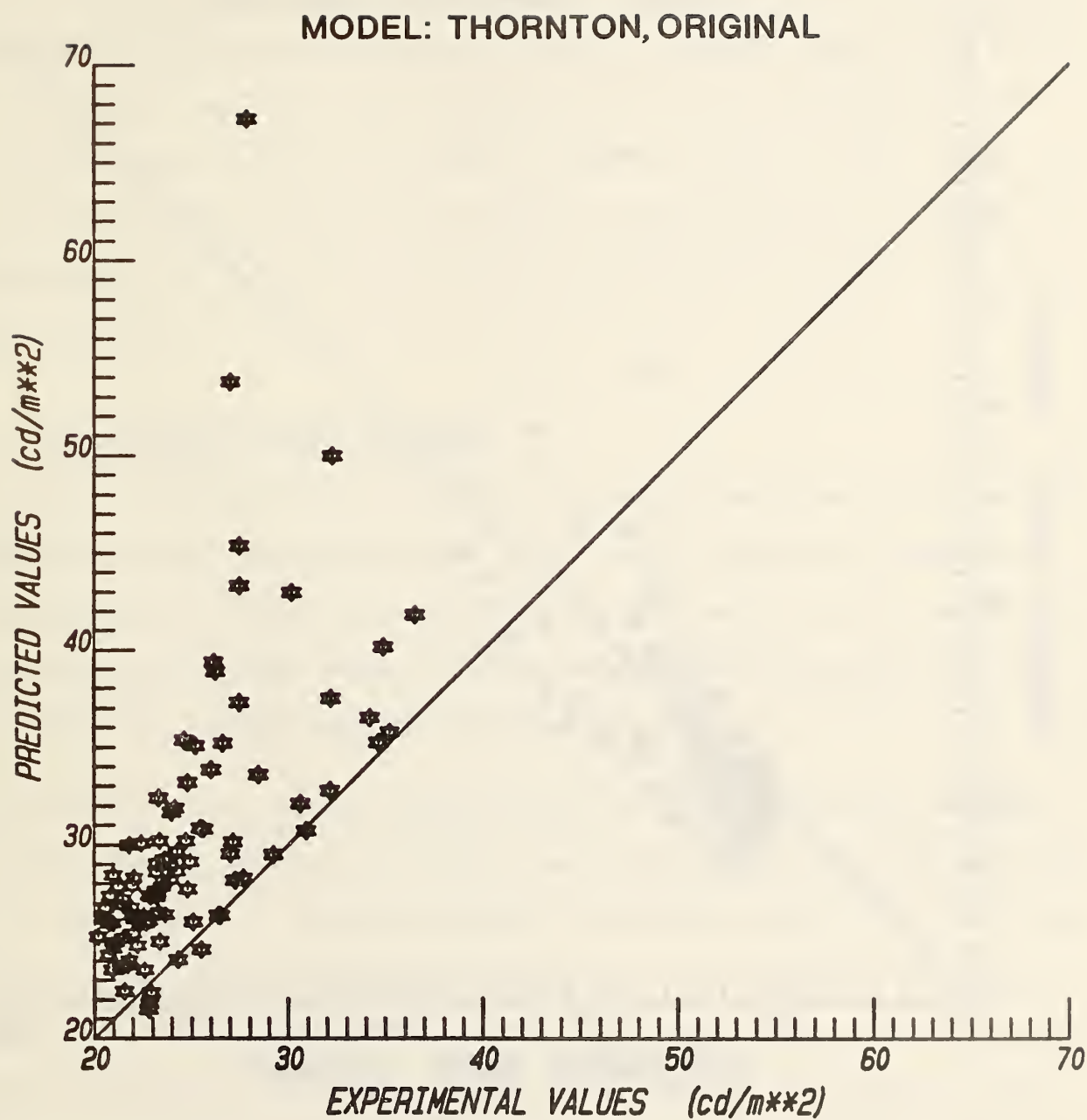


Figure 5. Correlation plot (scatter diagram) for original Thornton model. Predicted equivalent luminance is plotted against the equivalent luminance measured in the Sanders-Wyszecki (1964) experiment.

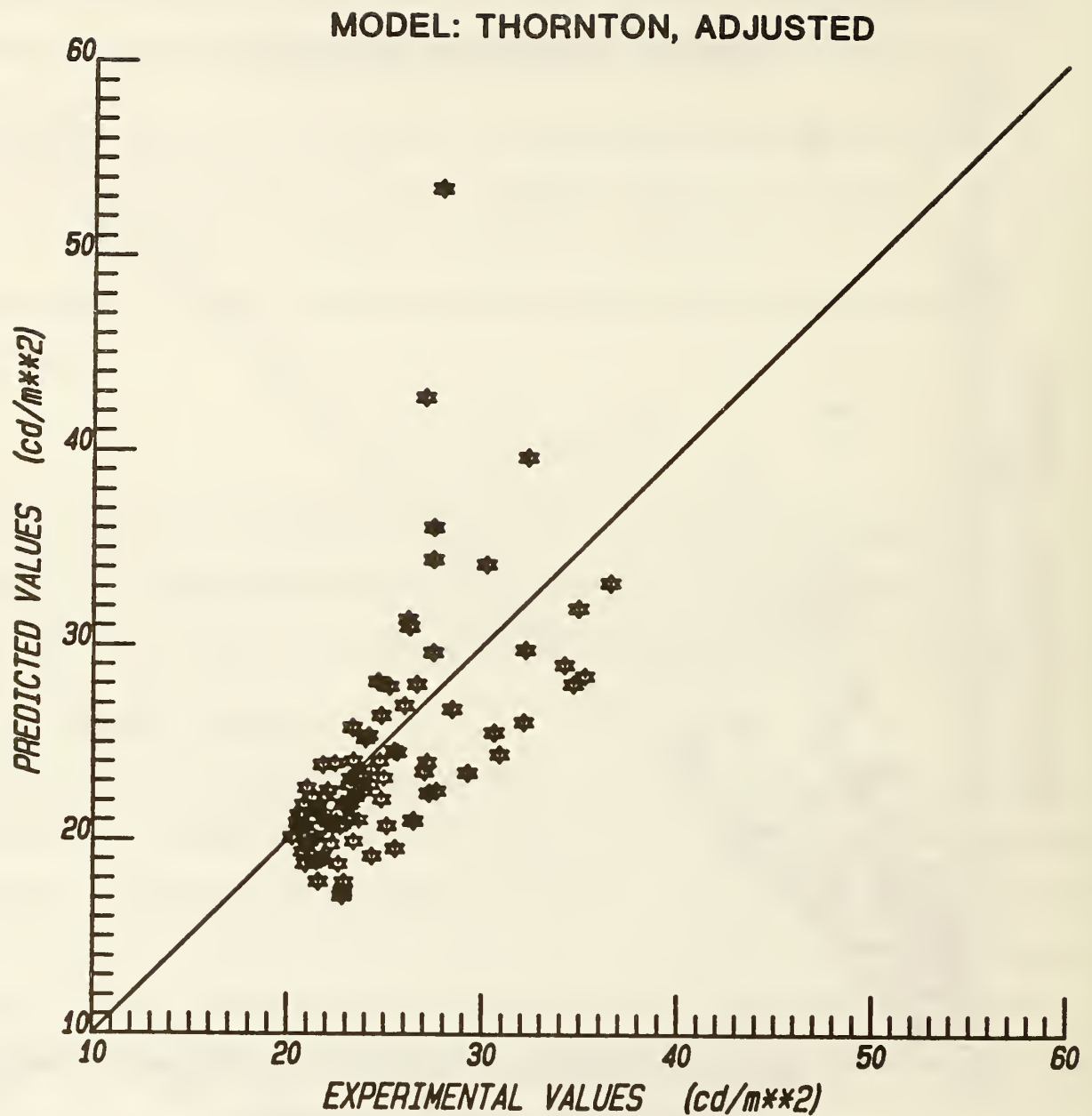


Figure 6. Correlation plot (scatter diagram) for adjusted Thornton model. Predicted equivalent luminance is plotted against the equivalent luminance measured in the Sanders-Wyszecki (1964) experiment. The absolute scale of the Thornton predictions (Fig. 5) was adjusted for optimum fit.

The correlation coefficient is 0.616 in both cases. It is worth recalling that the sums of squares associated with the author's optimized models are 36 (figure 1) for the very best model and 44 (figure 2) for the restricted model. Thus, the sum of squared errors of the author's restricted model is better than that associated with the original Guth and Thornton models by factors of 56 and 131, respectively; and better than even the adjusted models by factors of 12 and 44, respectively.

6.2 OPPONENT-LEVEL CHANNEL RESPONSES

The remainder of section 6 is devoted to showing the predictions of the various models for several basic functions of color vision. A model that predicts only brightness is valuable (especially in the lighting field), but a model that also predicts other aspects of color perception with reasonable accuracy is valued more highly by many vision specialists.

The figures shown as part of this subsection (6.2) are not primarily intended as predictions of empirical vision data, but represent the theoretical responses of the opponent-colors channels implied by the various models. These functions can in turn be used to predict experimental color data of various kinds. Guth believes that the two opponent curves (T and D) should be interpretable as direct predictions of the perceived hues of the spectrum (red-green and blue-yellow, respectively). If that assumption is accepted, then the figures of this subsection can also be interpreted as predictions of one type of color data.

6.2.1 Guth and Lodge

Guth's channel responses [Eqs. (1)], shown in figure 7, are considered reasonable in general form by most theorists. The luminance function, $\underline{A}$, is all-positive and very similar to the standard $V(\lambda)$ function. If these functions are interpreted (as Guth does) as directly reflecting the perceived colors of the pure (monochromatic) lights of the spectrum, this model fails to reflect the fact that redness reappears at the shortwave end of the spectrum (violet hue). There should be a small positive lobe of the red-green (dashed) curve in the wavelength region from 400 up to perhaps 450 nm or higher.

It should be noted that most of the color-related functions that can be predicted from the opponent-level channel responses do not depend on the absolute sizes of the coefficients, but only on various ratios that can be formed from them. Consequently, no further distinction will be made in this paper between the original and adjusted models of Guth and Lodge (see section 6.1.4) and Thornton (see section 6.1.6). Each subsection from here on will discuss the predictions of four models only: Howett, Best-Fitting; Howett, Restricted; Guth and Lodge (Original, but not so labeled); and Thornton (Original, but not so labeled).

6.2.2 Thornton

Thornton's model [Eqs. (3)], depicted in figure 8, does predict the redness of shortwave light. Its most unsatisfying feature is the double-humped yellow-blue function (which virtually coincides with the red-green function above about 540 nm). Note that Guth (figure 7) chose blue as the positive valence of his blue-yellow function. The choice of sign is arbitrary and the dotted curve in figure 7 can be mentally reflected in the abscissa axis for direct comparison with the other models.

MODEL: GUTH & LODGE

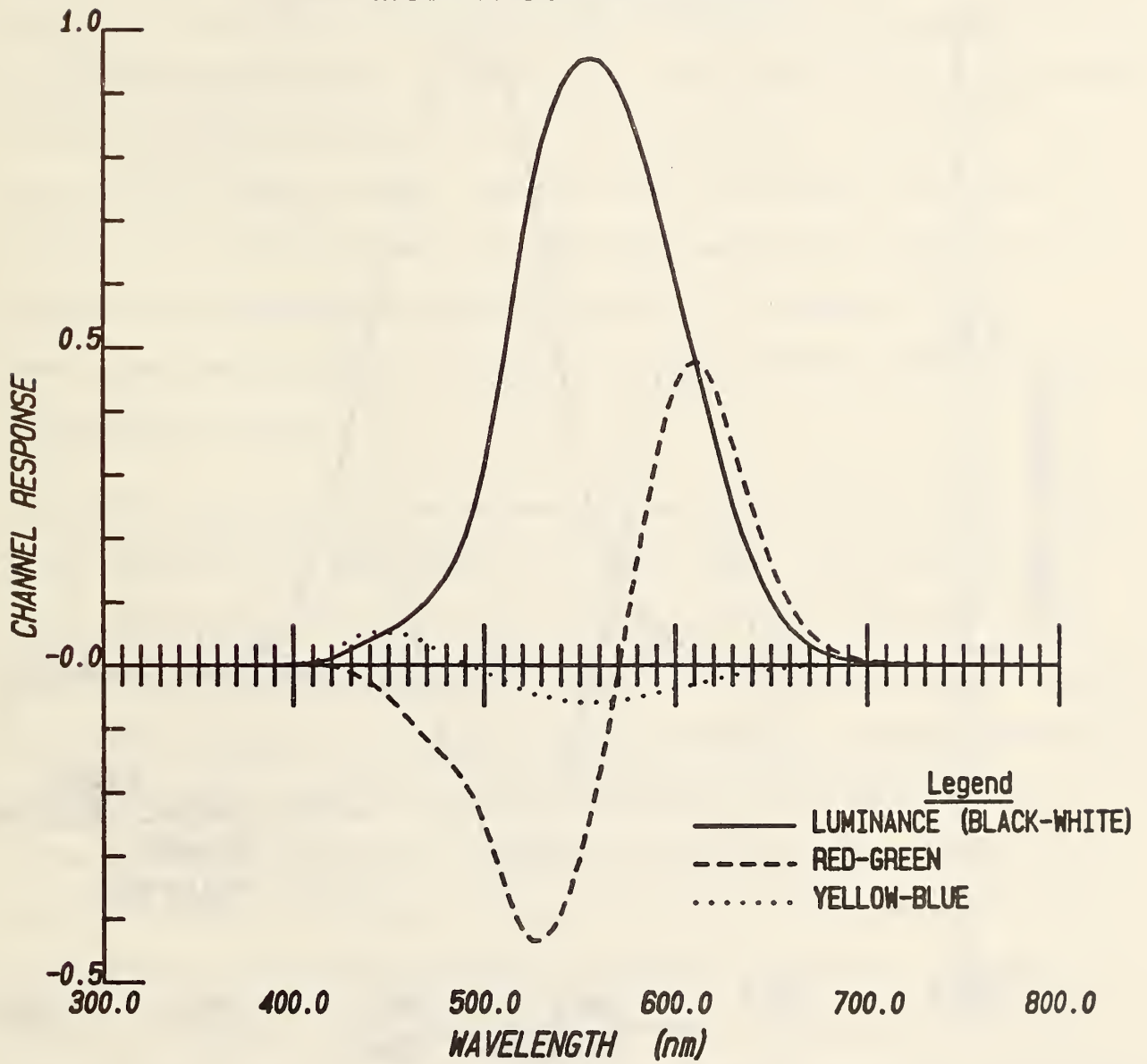


Figure 7. Opponent-level channel responses for the Guth and Lodge model.

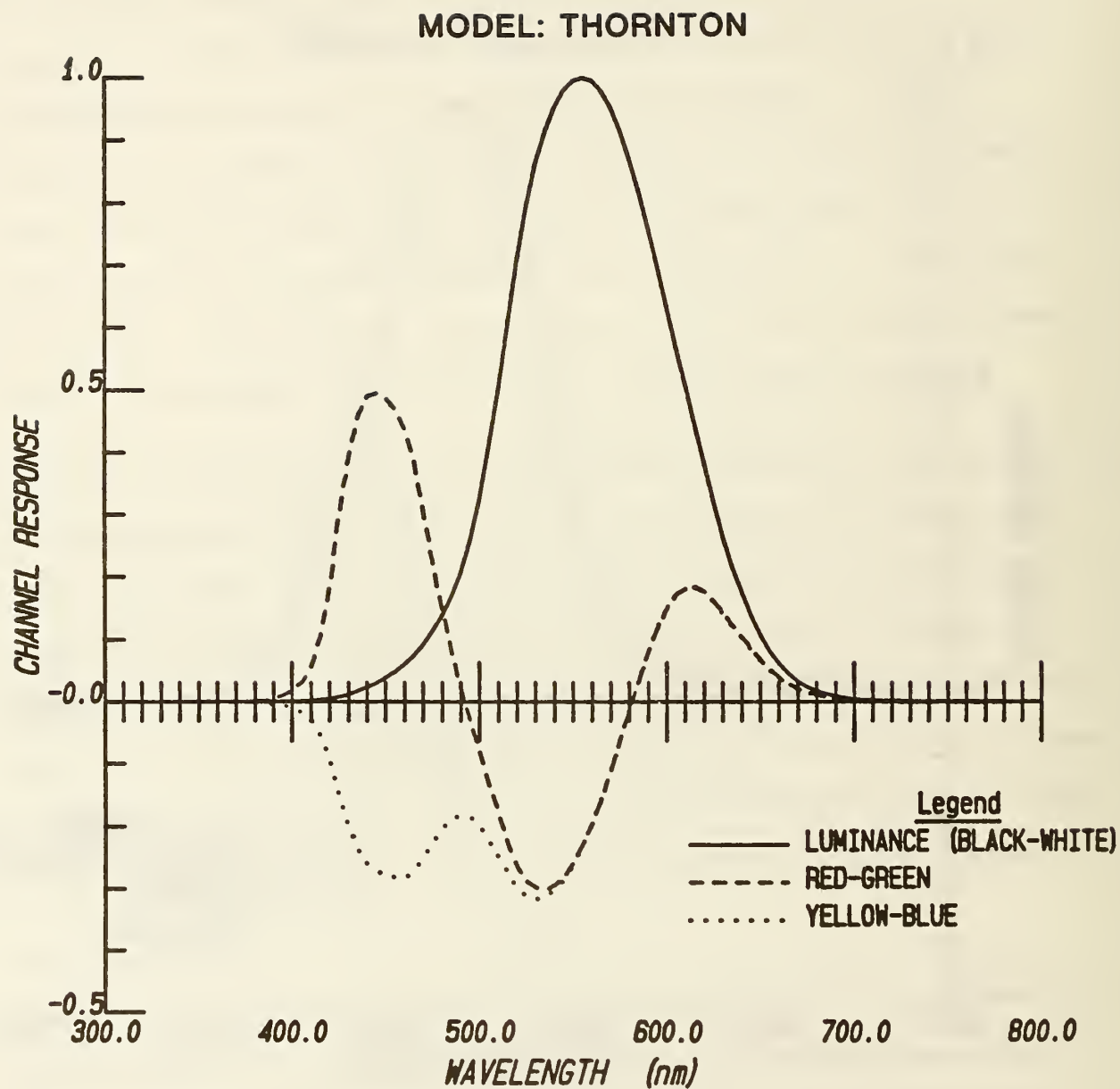


Figure 8. Opponent-level channel responses for the Thornton model.

6.2.3 Howett, Best-Fitting

The functions shown in figure 9 characterize the best unrestricted model optimized by the author [Eqs. (8)]. This model was derived entirely on the basis of the Sanders and Wyszecki brightness data. There were no restrictions imposed to make the model meaningful for other color predictions. Nevertheless, the forms of these channel-response functions bear an encouraging resemblance to the Guth and the Thornton functions, which were restricted to make sense with respect to color predictions (especially Guth's). The luminance function looks a good deal like $V(\lambda)$, as it should, but the shortwave negative lobe is an unrealistic feature.

It should be mentioned that the Sanders and Wyszecki (1964) experiment involved a large field of view, subtending 10° of visual angle. The chromaticity coordinates given for the test colors in their paper were calculated on the basis of the CIE 1964 supplementary 10° observer, rather than the more commonly used 1931 2° observer. The a_{ij} coefficients that emerged from the optimizations [Eqs. (8) and (9)] accordingly apply to tristimulus values within the 10° system. However, the 10° and 2° observers, while differing significantly, do not differ greatly. Therefore, the color functions predicted by the author's two models were all constructed using the spectral tristimulus values of the CIE 1931 2° observer, rather than the 10° observer to which the coefficients actually apply. This was done to permit direct comparison with the predictions of the Guth model and the Thornton model. Like the use of Guth's formulas for predicting the various color functions (see section 6), the switch of colorimetric systems provided a further handicap to the author's models, so that a reasonably good showing relative to the other models might be taken as providing hope

MODEL: HOWETT, BEST FITTING

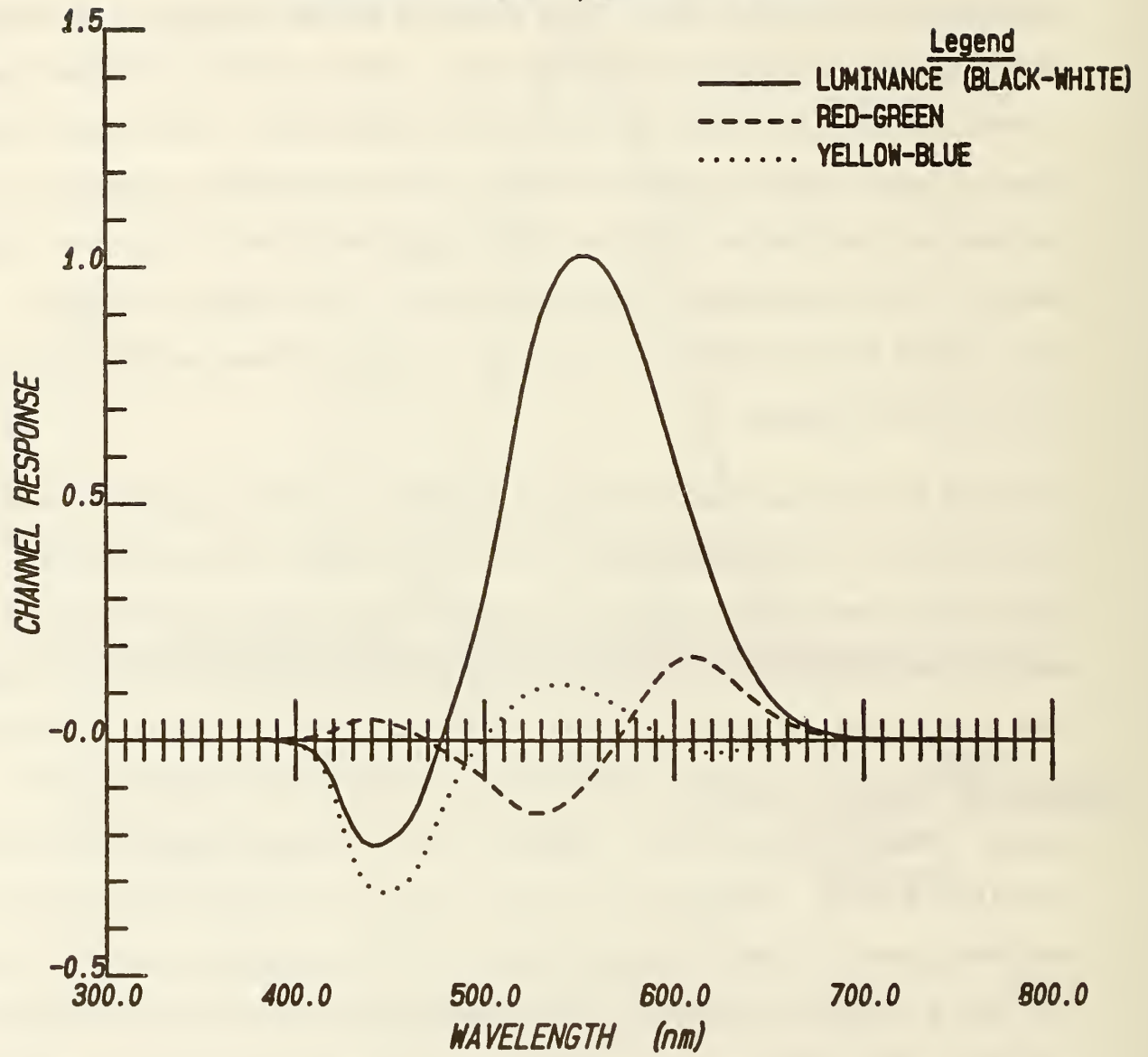


Figure 9. Opponent-level channel responses for the best-fitting Howett model.

that appropriate further adjustments would improve the predictions. Conducting this study with this level of imprecision also weighed against the application of any statistical tests at this stage (see the end of section 5), although it should be kept in mind that the prediction of equivalent luminances by the author's models was not subject to the handicaps that have been mentioned.

6.2.4 Howett, Restricted

The author's model was re-optimized with the single restriction that the luminance curve be all-positive [Eqs. (9)]. Figure 10 now shows curves that make about as much sense as Guth's or Thornton's curves. The reddish appearance of the shortwave band of the spectrum is predicted (small positive red lobe). However, there is a secondary negative (blue) lobe of the yellow-blue function in the range of 590-700 nm, which does not correspond to hue perceptions in that region. The same feature appeared in the wholly unrestricted model (Fig. 9), and seems to be forced by the data.

Possibly, the interpretation that the opponent functions T and D directly specify the perceived hues of the spectrum, in simple red-green/blue-yellow terms, is an oversimplification. Perhaps another mathematical stage, of other than purely linear form, must be introduced to derive predictions of the hues of the spectrum from the opponent-channel curves.

6.3 SATURATION OF THE SPECTRUM

The curves shown in this subsection predict the saturation (perceived degree of difference from white) of spectrally pure lights. This function has been measured experimentally many times, and the details of the curve vary from one determination to another. The general form found is characterized by a minimum

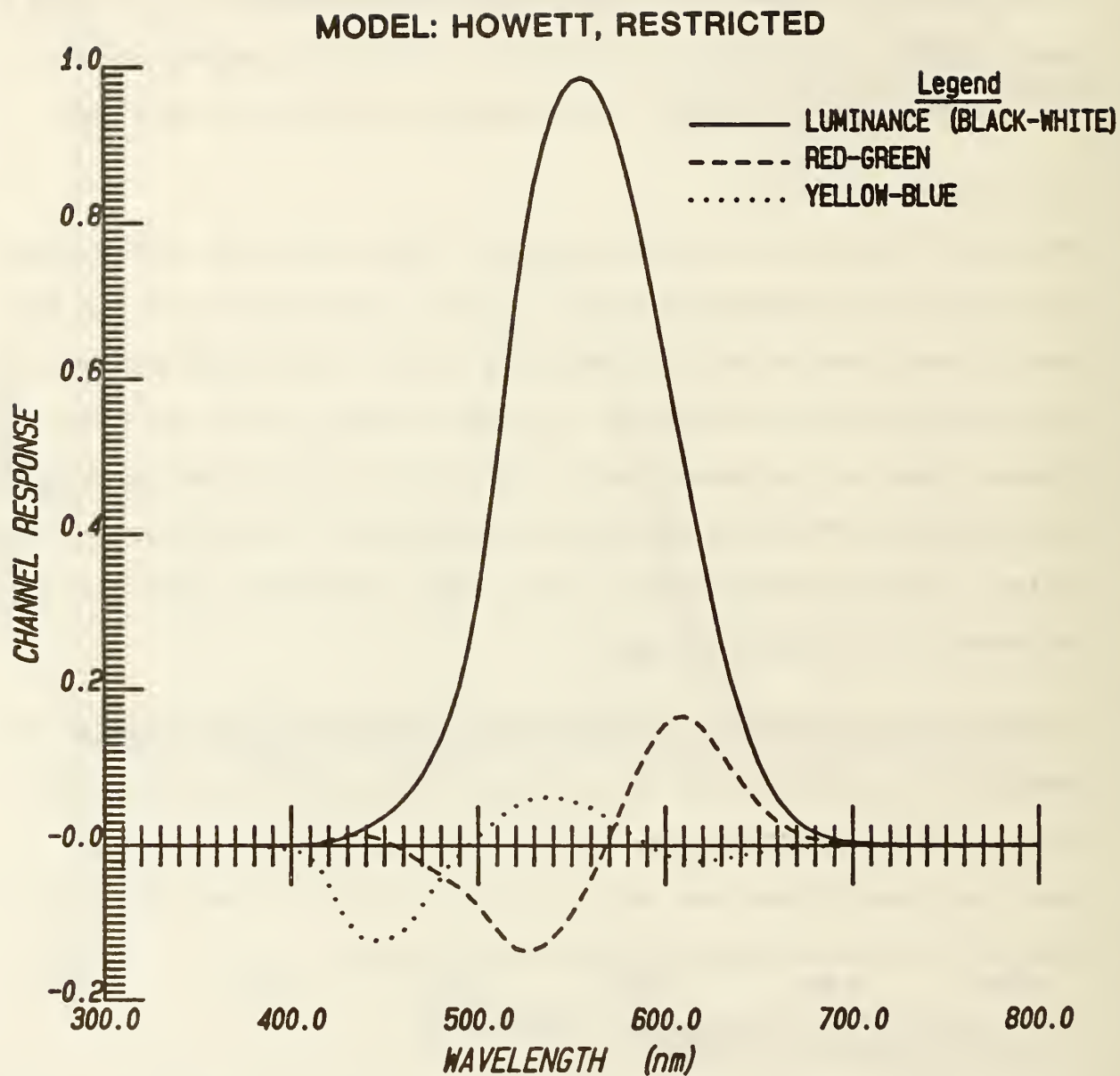


Figure 10. Opponent-level channel responses for the restricted Howett model.

saturation in the yellow, around 570-580 nm, and maximum saturation at both extremes of the spectrum (violet/blue and red). All the models correctly predict these gross features.

6.3.1 Guth and Lodge

Guth's model generates a saturation curve (figure 11) that probably resembles the typical experimental determinations best. Note the bump near 480 or 490 nm.

The formula used by Guth for predicting the saturation of any color is

$$\frac{|T|+|D|}{A+|T|+|D|} \quad (11)$$

The luminance or whiteness, A, in the denominator of Eq. (11) does not require an absolute value sign since it is never negative. The formula (11) has been used for all the saturation predictions in this subsection.

Intuitively, formula (11) defines saturation as the fractional part of the total absolute stimulation of the three opponent-level channels that arises from the two chromatic (opponent) channels. Presumably, Guth tried several different formulas for saturation and found that the expression (11) gave the most realistic-looking function. However, the use of the simple sums of absolute values appears more appropriate to Thornton's model than to Guth's. For the general linear model, it would also seem appropriate to test the expression

$$\frac{(|T|^p+|D|^p)^{1/p}}{(A^p+|T|^p+|D|^p)^{1/p}} \quad (12)$$

where the value of p associated with the particular model is used. The author has not yet examined any predictive function other than formula (11).

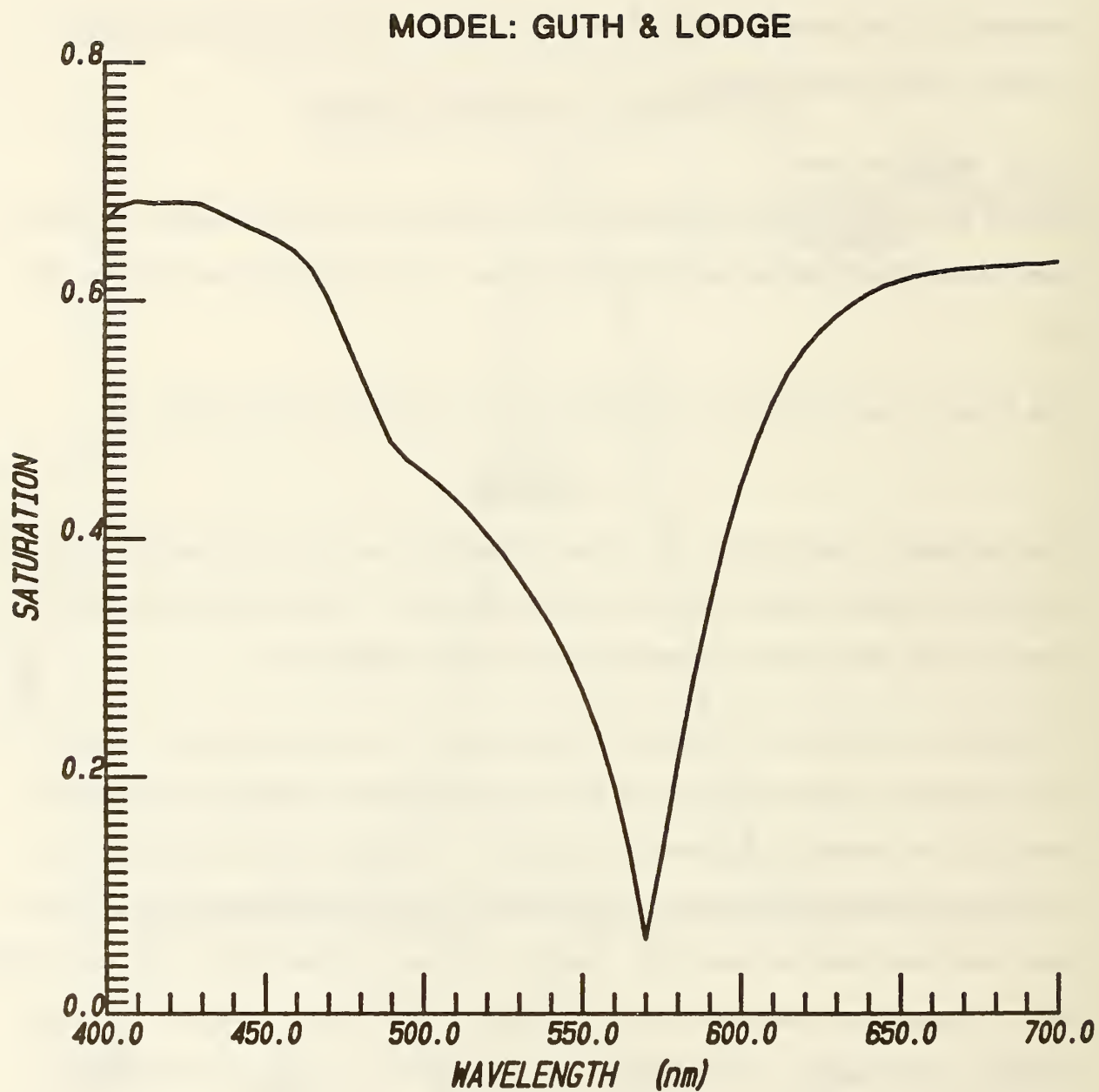


Figure 11. Saturation (perceived color strength) of the monochromatic lights of the visible spectrum, as predicted by the Guth and Lodge model.

It is interesting that in the very latest model to which Guth has contributed (Benzschawel and Guth, 1984), saturation is no longer predicted by the expression (11), but rather by (12) with $p=2$, the characteristic value of Guth's vector model. This latest model, which is nonlinear, is intended primarily to predict color differences, and it is not yet clear whether Guth regards it as a general color-vision model.

6.3.2 Thornton

Thornton's spectral saturation function, shown in figure 12, exaggerates the bump barely seen in Guth's curve (figure 11). Many functions of color vision have minor bumps present, which often become considerably more conspicuous under conditions of chromatic adaptation or adaptation to high levels of white light. See Guth, Massof, and Benzschawel (1980) for data reflecting this general phenomenon.

6.3.3 Howett, Best-Fitting

It is seen in figure 13 that the author's unrestricted model implies a saturation function generally resembling Thornton's (figure 12). The high spike around 470 nm is unrealistic, and is presumably a consequence of the negative lobe of the luminance function (figure 9).

6.3.4 Howett, Restricted

The author's restricted model generates a saturation function (figure 14) that has lost the incorrect spike (figure 13) and now bears considerable resemblance to Thornton's curve (figure 12).

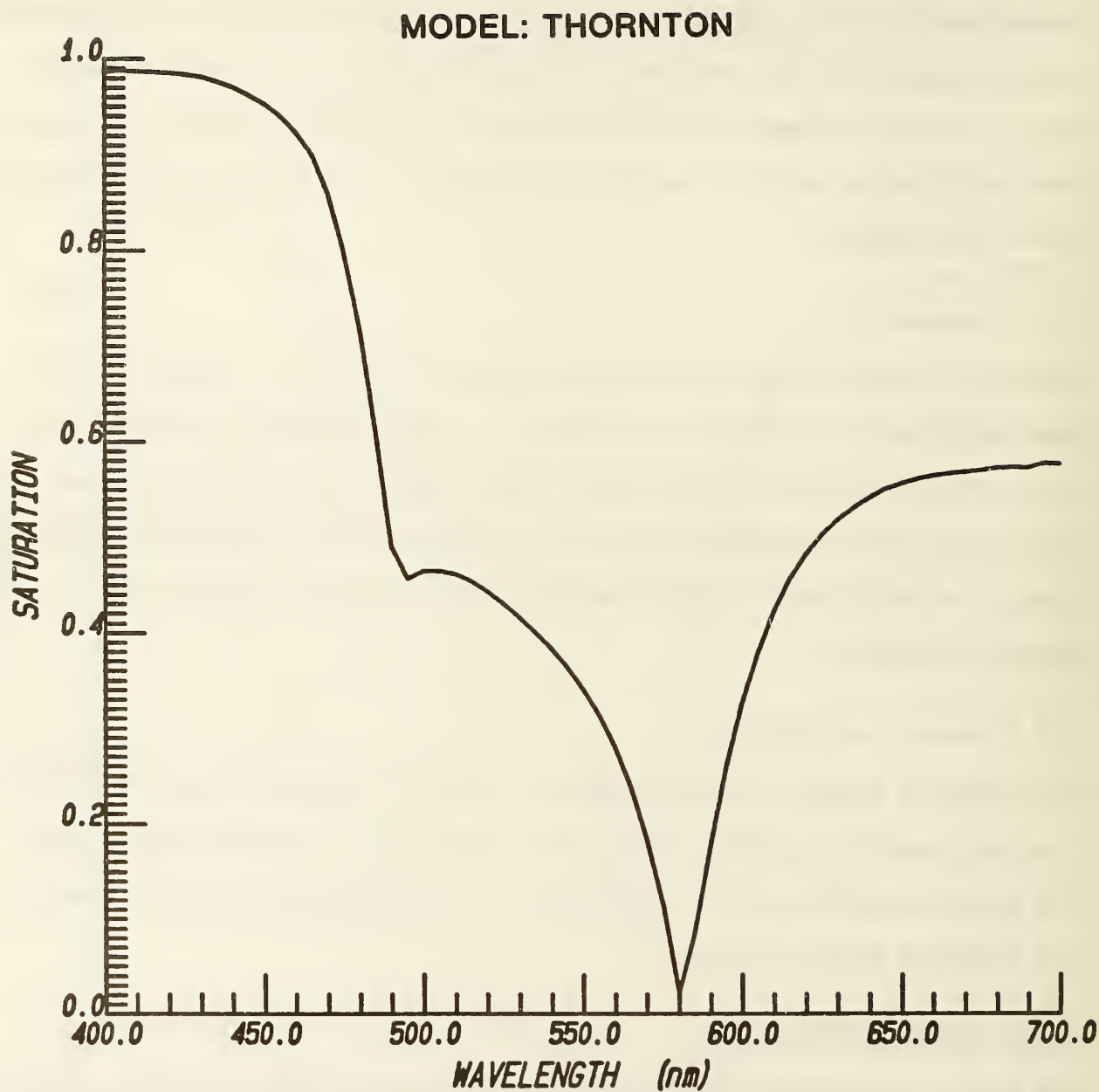


Figure 12. Saturation (perceived color strength) of the monochromatic lights of the visible spectrum, as predicted by the Thornton model.

MODEL: HOWETT, BEST-FITTING

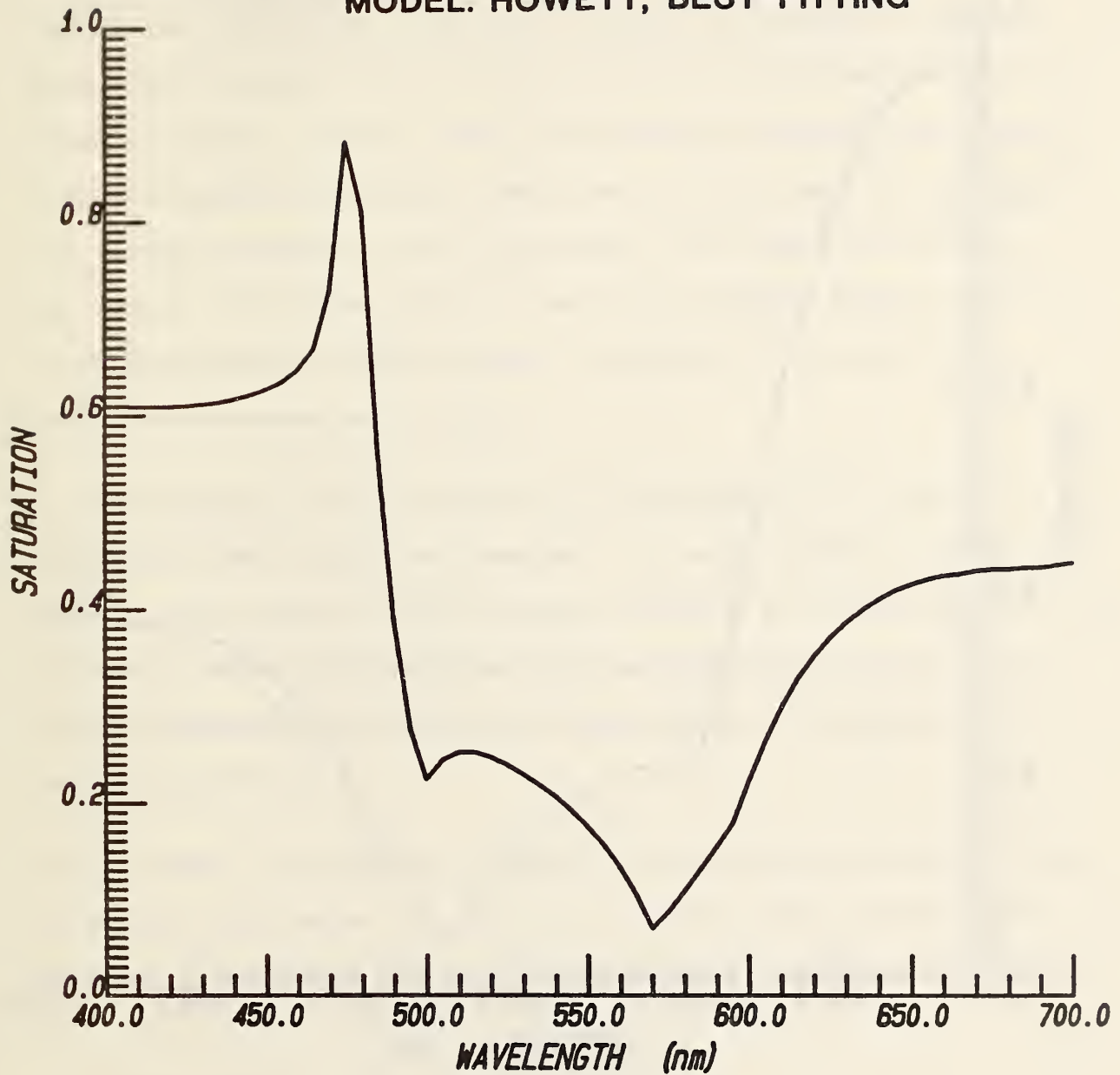


Figure 13. Saturation (perceived color strength) of the monochromatic lights of the visible spectrum, as predicted by the best-fitting Howett model.

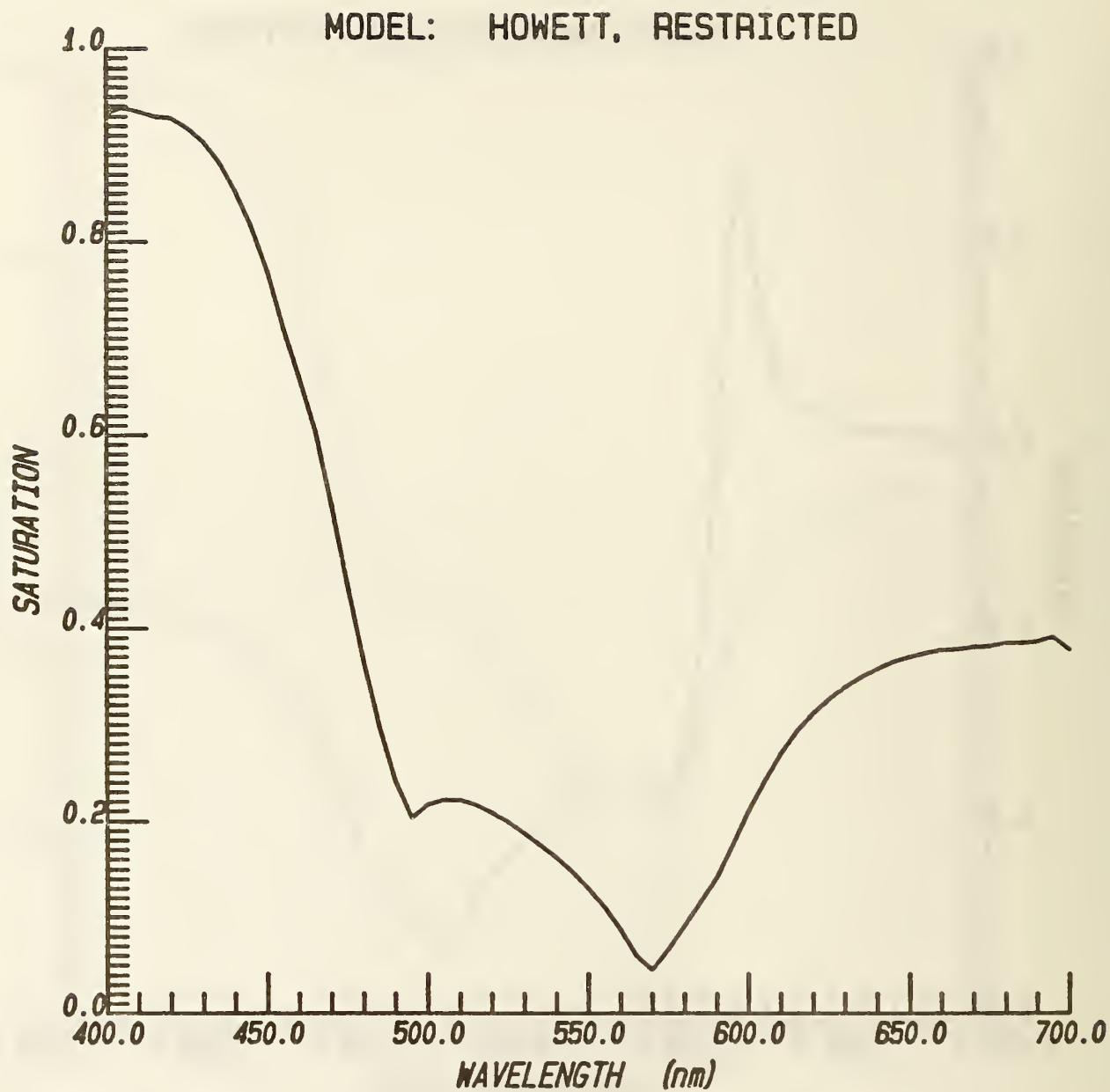


Figure 14. Saturation (perceived color strength) of the monochromatic lights of the visible spectrum, as predicted by the restricted Howett model.

6.4 EQUIVALENT LUMINANCE OF THE SPECTRUM

When the overall spectral sensitivity of the eye is determined by flicker photometry or certain other procedures, the result is the standard CIE $V(\lambda)$ luminous efficiency function. When the criterion is changed to an appearance of equal brightness (or certain other procedures), the sensitivity function that results is similar to $V(\lambda)$, but broader, i.e., higher at the ends of the spectrum. The latter curve also tends to be somewhat bumpier than $V(\lambda)$. CIE Publication No. 41 (1978) provides a review of the two classes of sensitivity-determination techniques.

In this subsection, the predictions by the four models of the latter (brightness-based) function are examined. The author (Howett, in press) refers to this function as "the equivalent luminance of the spectrum," or "the equivalent luminous efficiency function," and symbolizes it as $V_q(\lambda)$. (No standard terminology or symbol has yet been adopted by an international or national organization.)

For each model, the equivalent luminance of the spectrum is calculated using the spectral tristimulus form of Eq. (6), with the channel responses [Eqs. (5)] and power, p , associated with that particular model. Explicitly, the equation is

$$V_q(\lambda) = \{ |\bar{a}(\lambda)|^p + |\bar{t}(\lambda)|^p + |\bar{d}(\lambda)|^p \}^{1/p} . \quad (13)$$

The function $V_q(\lambda)$ specifies the normalized (relative) equivalent luminance of each wavelength in an equal-energy spectrum, just as $V(\lambda)$ corresponds to the normalized luminance of each wavelength in an equal-energy spectrum. "Normalized"

in the preceding sentence (a more specific term than the more commonly used "relative") means that all values along the curve were divided by the maximum value, so that the maximum value of the normalized curve becomes unity.

6.4.1 Guth and Lodge

The equal-brightness "sensitivity" curve implied by the Guth model is shown as the solid curve in figure 15. (This is a prediction of the curve that has been referred to formally in this paper as the "equivalent luminance of the spectrum," or the "equivalent luminous efficiency function.") Guth's curve resembles experimental data more than the others do. The dashed curve is CIE $V(\lambda)$. Both curves are normalized to unit maximum for comparison. Note the slight bump in the neighborhood of 600 nm (here only an inflection).

6.4.2 Thornton

Thornton's corresponding function (the solid curve in figure 16) has a more conspicuous bump near 600 nm, and a very large bump --which is probably unrealistic-- around 450 nm.

6.4.3 Howett, Best-Fitting

The author's unrestricted model, shown as the solid curve in figure 17, predicts an equivalent-luminance curve that strikingly resembles the Thornton prediction (figure 16). Again, the large bump near 450 nm is unrealistic, and presumably reflects the negative lobe in the luminance channel response (figure 9).

6.4.4 Howett, Restricted

The author's restricted model generates a curve (figure 18) in which the bump at 450 nm is greatly suppressed, but still clearly present. The bumpiness near and below 600 nm is also still evident. Note in figure 15 that even though

MODEL: GUTH & LODGE

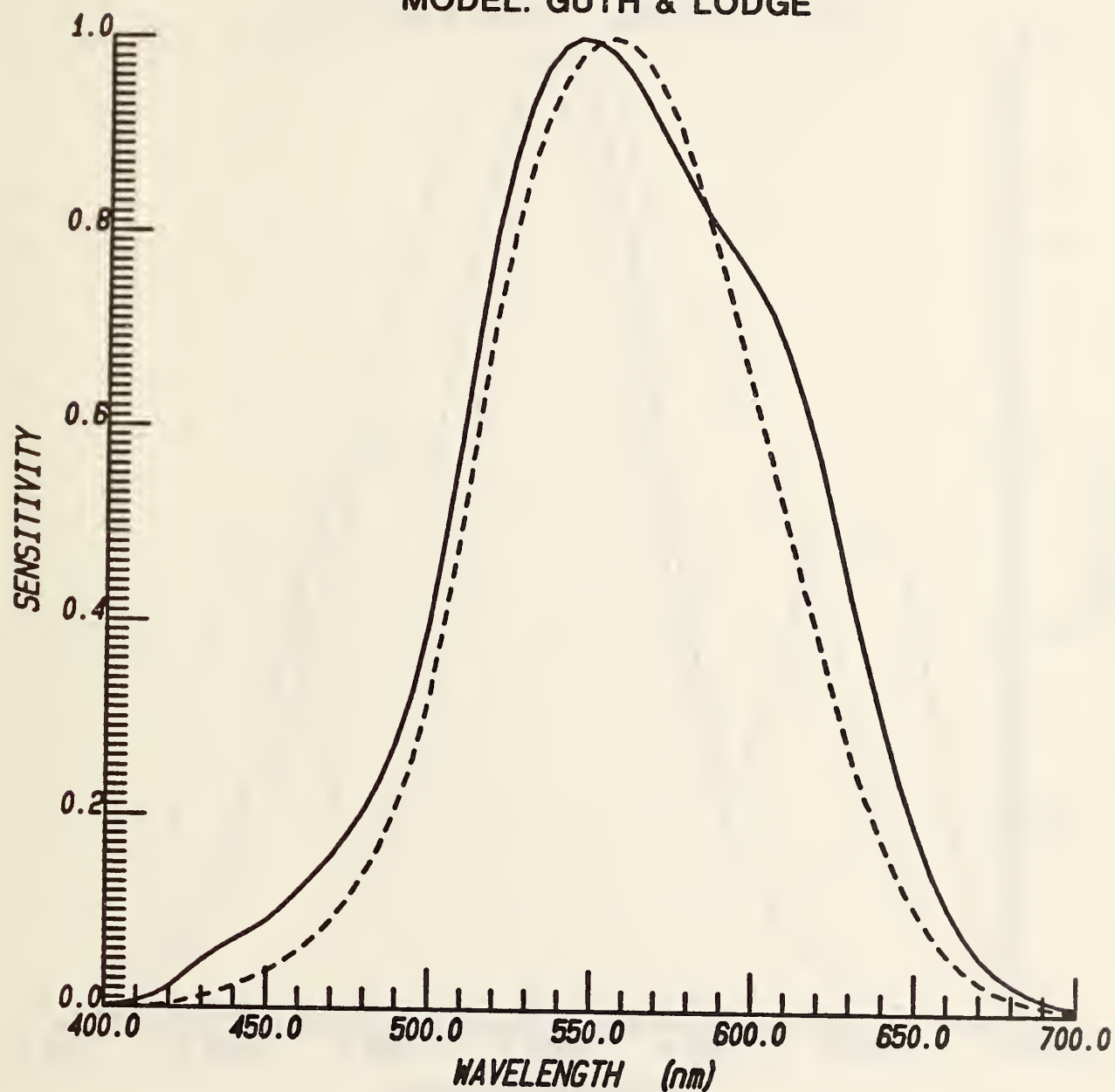


Figure 15. Equivalent (brightness-based) spectral luminous efficiency function (solid curve), as predicted by the Guth and Lodge model. The dashed curve is the standard (flicker-based) CIE luminous efficiency function, both curves being normalized to unit maximum.

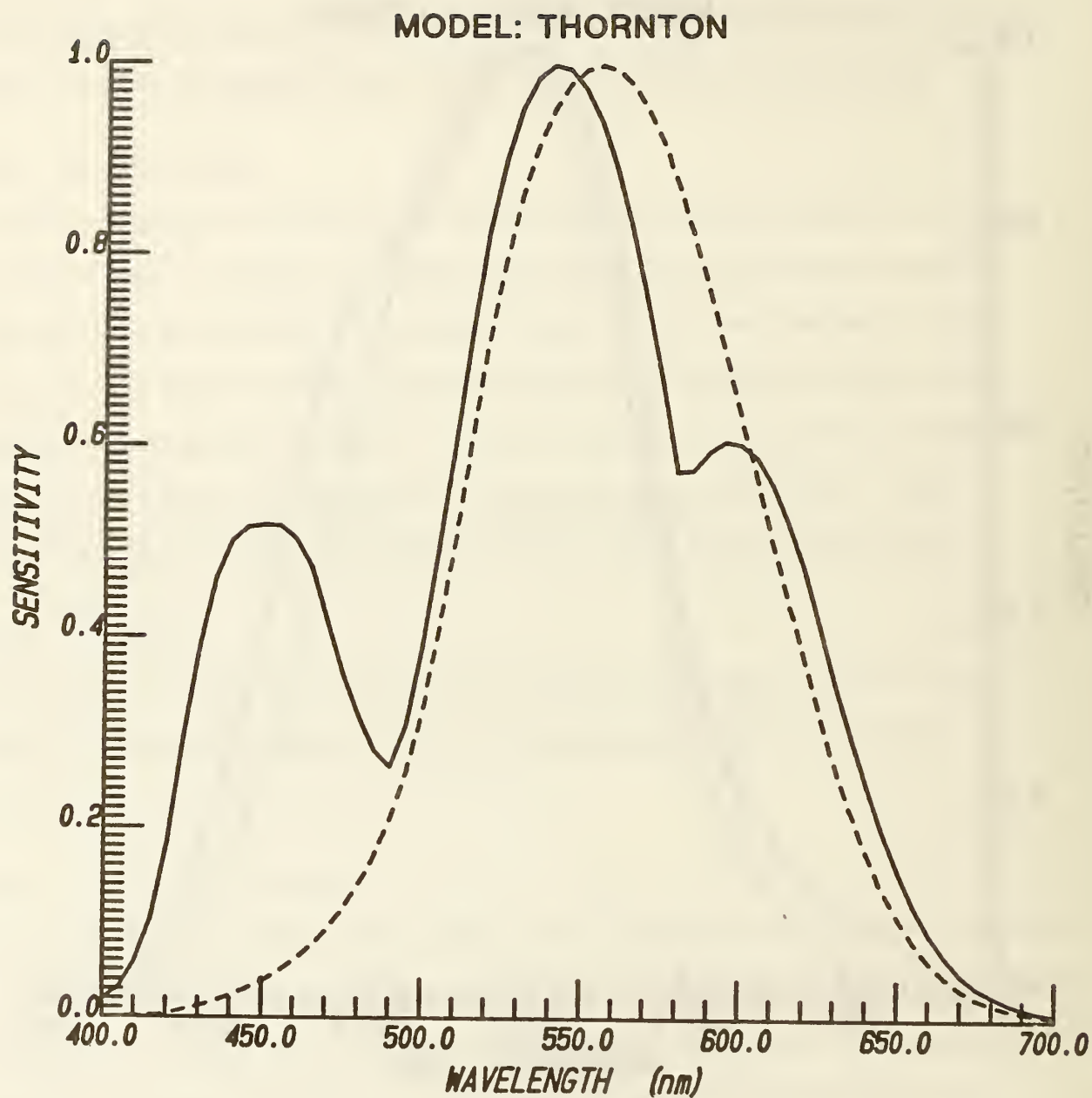


Figure 16. Equivalent (brightness-based) spectral luminous efficiency function (solid curve), as predicted by the Thornton model. The dashed curve is the standard (flicker-based) CIE luminous efficiency function, both curves being normalized to unit maximum.

MODEL: HOWETT, BEST-FITTING

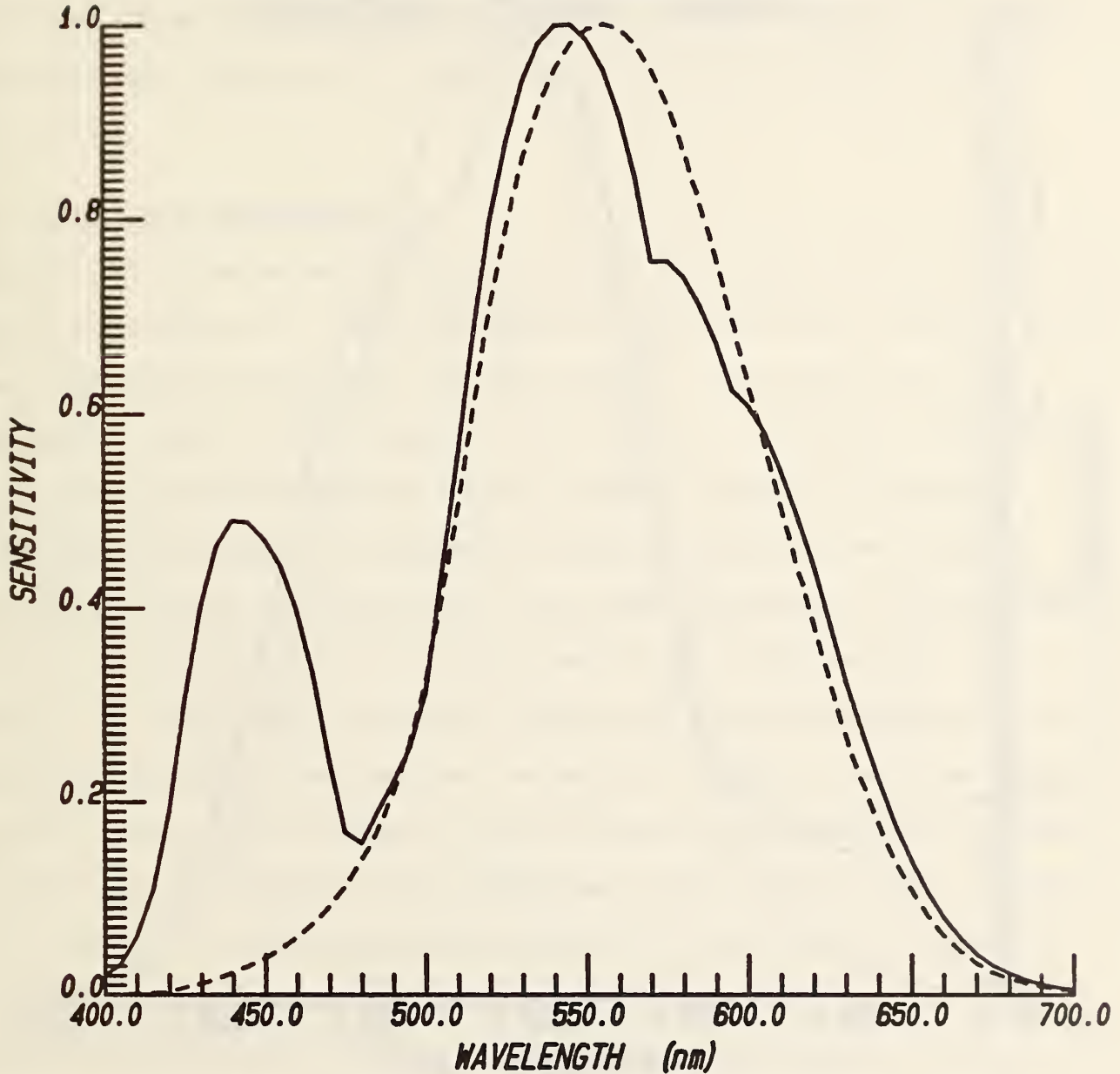


Figure 17. Equivalent (brightness-based) spectral luminous efficiency function (solid curve), as predicted by the best-fitting Howett model. The dashed curve is the standard (flicker-based) CIE luminous efficiency function, both curves being normalized to unit maximum.

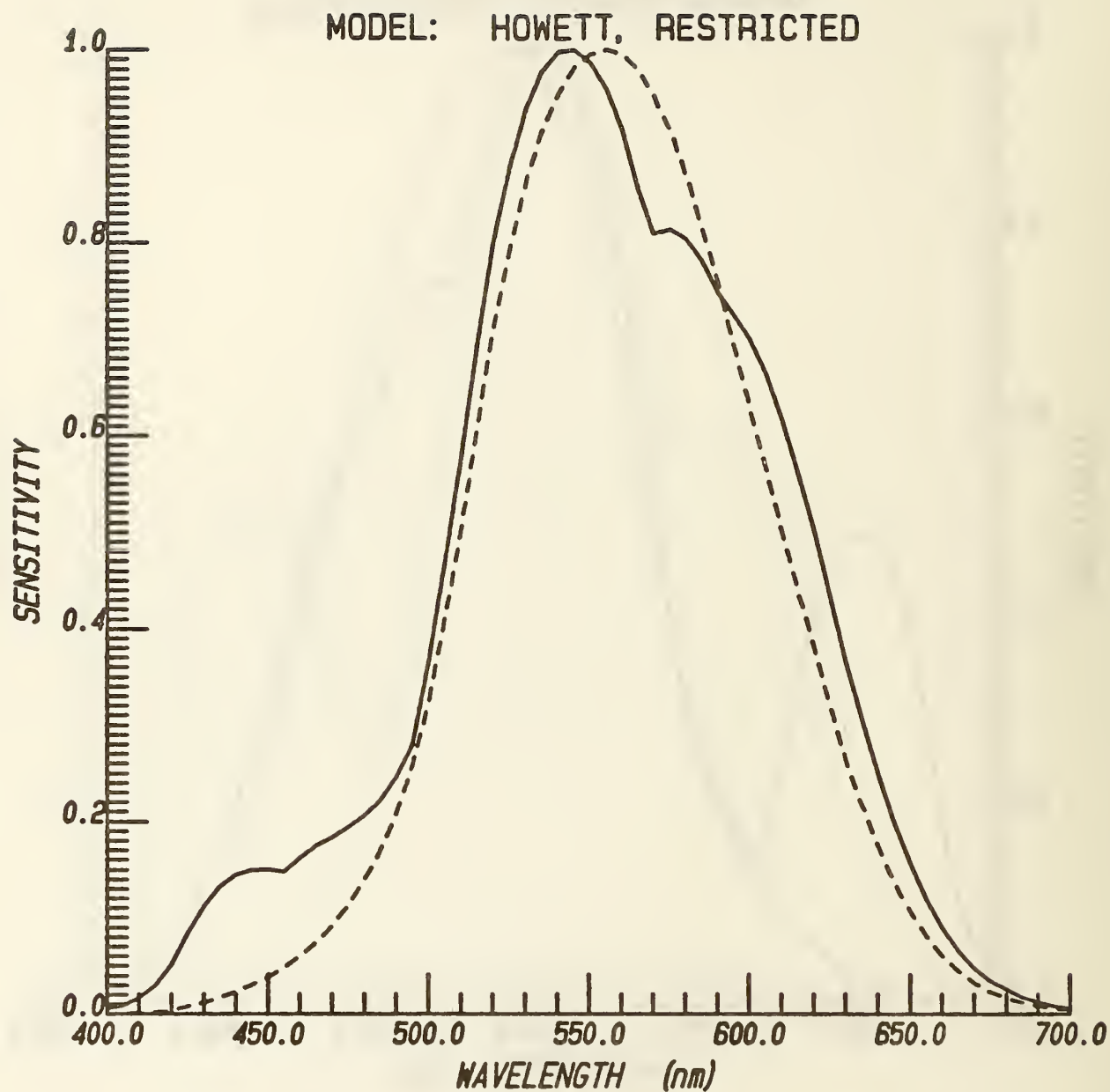


Figure 18. Equivalent (brightness-based) spectral luminous efficiency function (solid curve), as predicted by the restricted Howett model. The dashed curve is the standard (flicker-based) CIE luminous efficiency function, both curves being normalized to unit maximum.

Guth's predicted curve is much smoother, slight inflections suggesting the bump near 450 nm (as well as, more pronouncedly, that near 600 nm) are visible. Published data resemble Guth's smoother curve more than the bumpier curve of this figure.

6.5 WAVELENGTH DISCRIMINATION

The wavelength discrimination function is one of the classical curves of color vision experimentation. As with the spectral saturation curve, there have been many published determinations, with some variation in fine details but with a few features that are always present. The curve plots the amount of wavelength change that can just be visually detected (ordinate) against the starting wavelength as abscissa. The threshold is usually taken to be the average of the just noticeable wavelength shifts in the upward and downward directions from the starting wavelength (except near the ends of the visible spectrum). All four of the models being studied here predict wavelength discrimination curves having the requisite gross features: two minima, in the blue and yellow portions of the spectrum; a maximum in the green region; and strong deterioration of wavelength discrimination (high thresholds) toward both ends of the spectrum. It is difficult to judge which model predicts the curve most closely resembling empirical data, because the various experimental determinations differ so much from each other in their details. No detailed, separate comments on figures 19-22 will be made here, for that reason. It can be seen in figures 21 and 22 that both of the author's models generate slightly bumpier wavelength discrimination curves than do the Guth and Thornton models (Figs. 19 and 20). It is difficult to judge whether the various actual data curves do or do not exhibit these minor irregularities.

MODEL: GUTH & LODGE

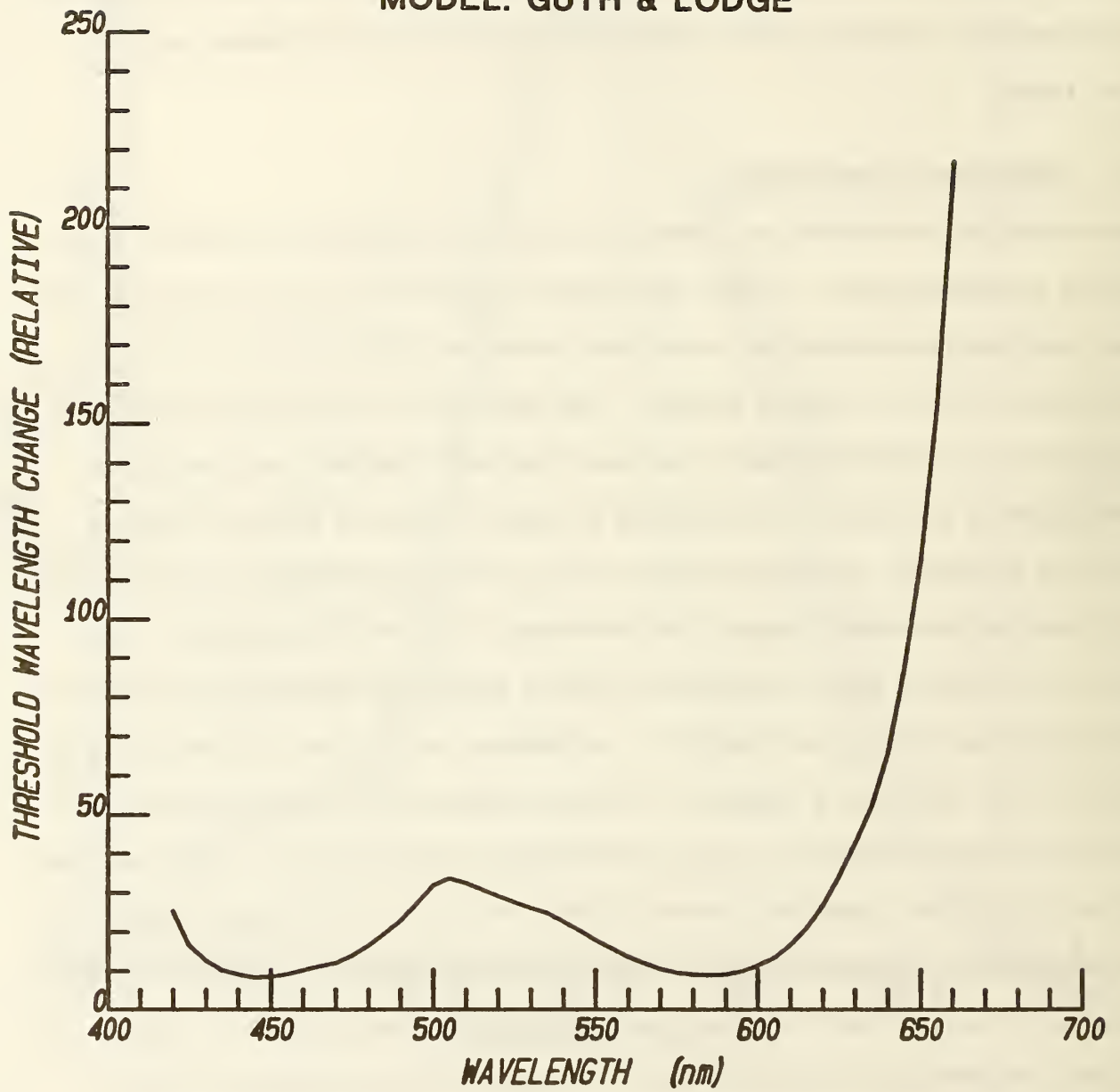


Figure 19. Wavelength discrimination function, as predicted by the Guth and Lodge model. The units on the ordinate axis are arbitrary.

MODEL: THORNTON

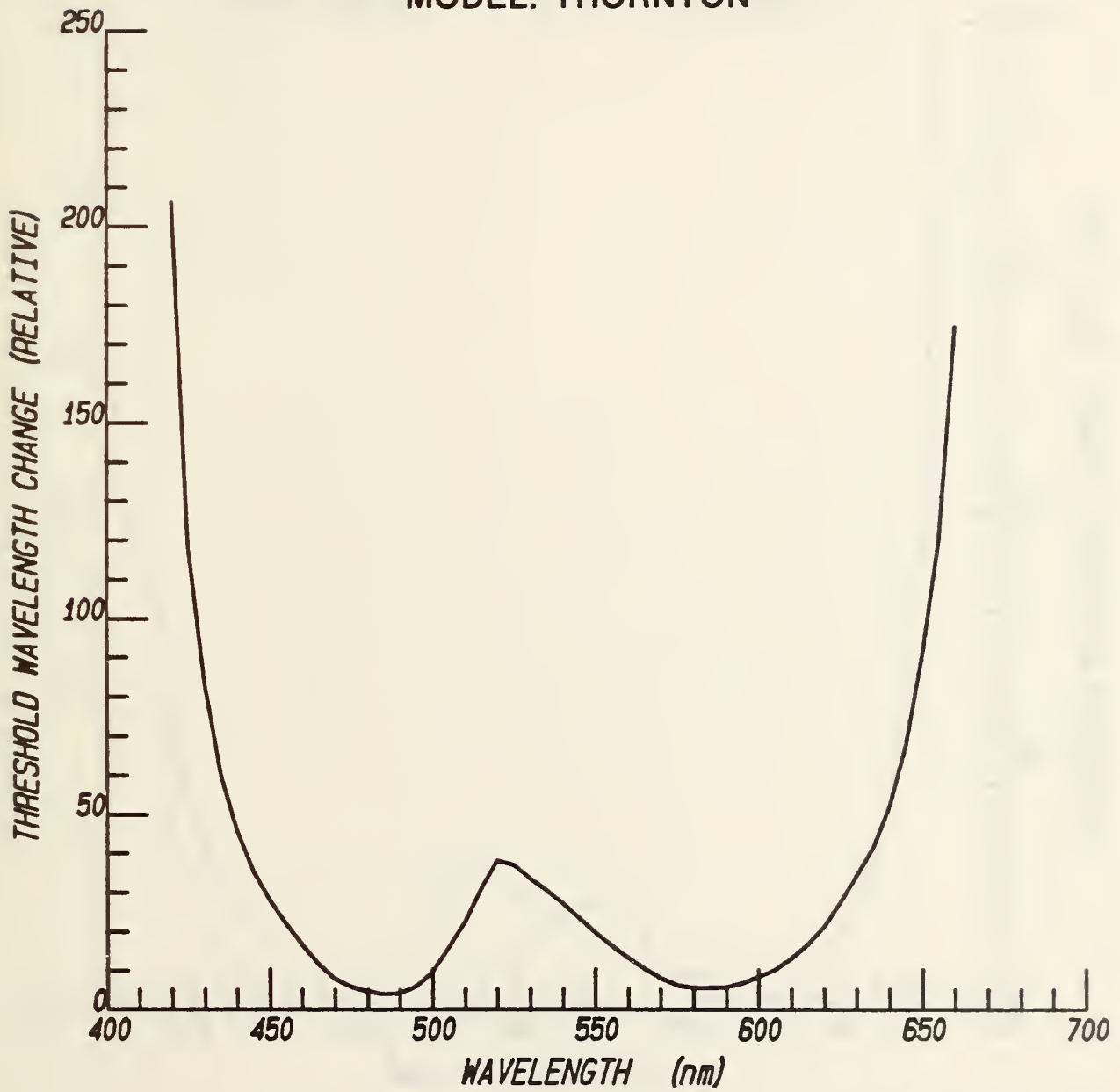


Figure 20. Wavelength discrimination function, as predicted by the Thornton model. The units on the ordinate axis are arbitrary.

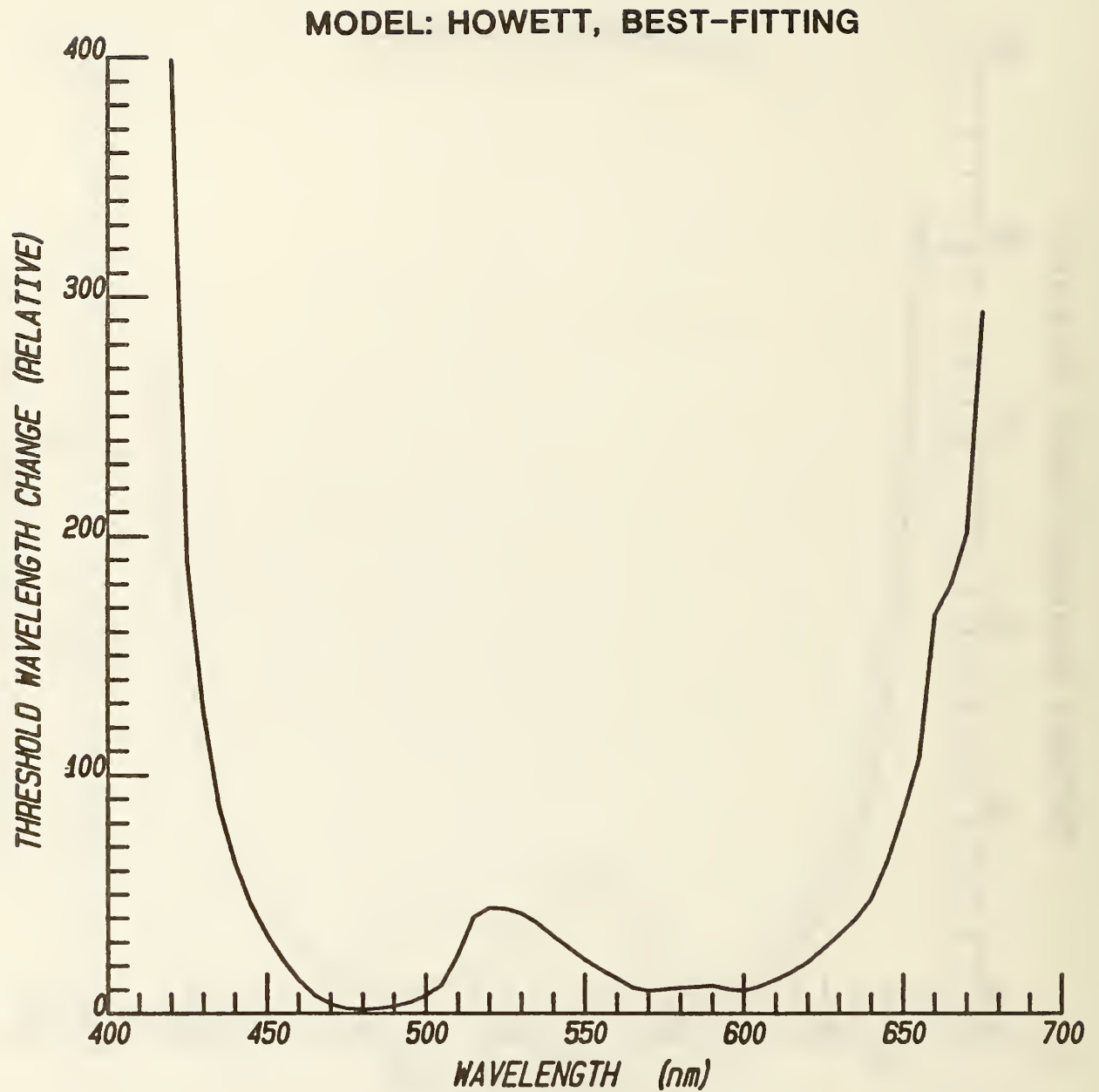


Figure 21. Wavelength discrimination function, as predicted by the best-fitting Howett model. The units on the ordinate axis are arbitrary.

MODEL: HOWETT, RESTRICTED

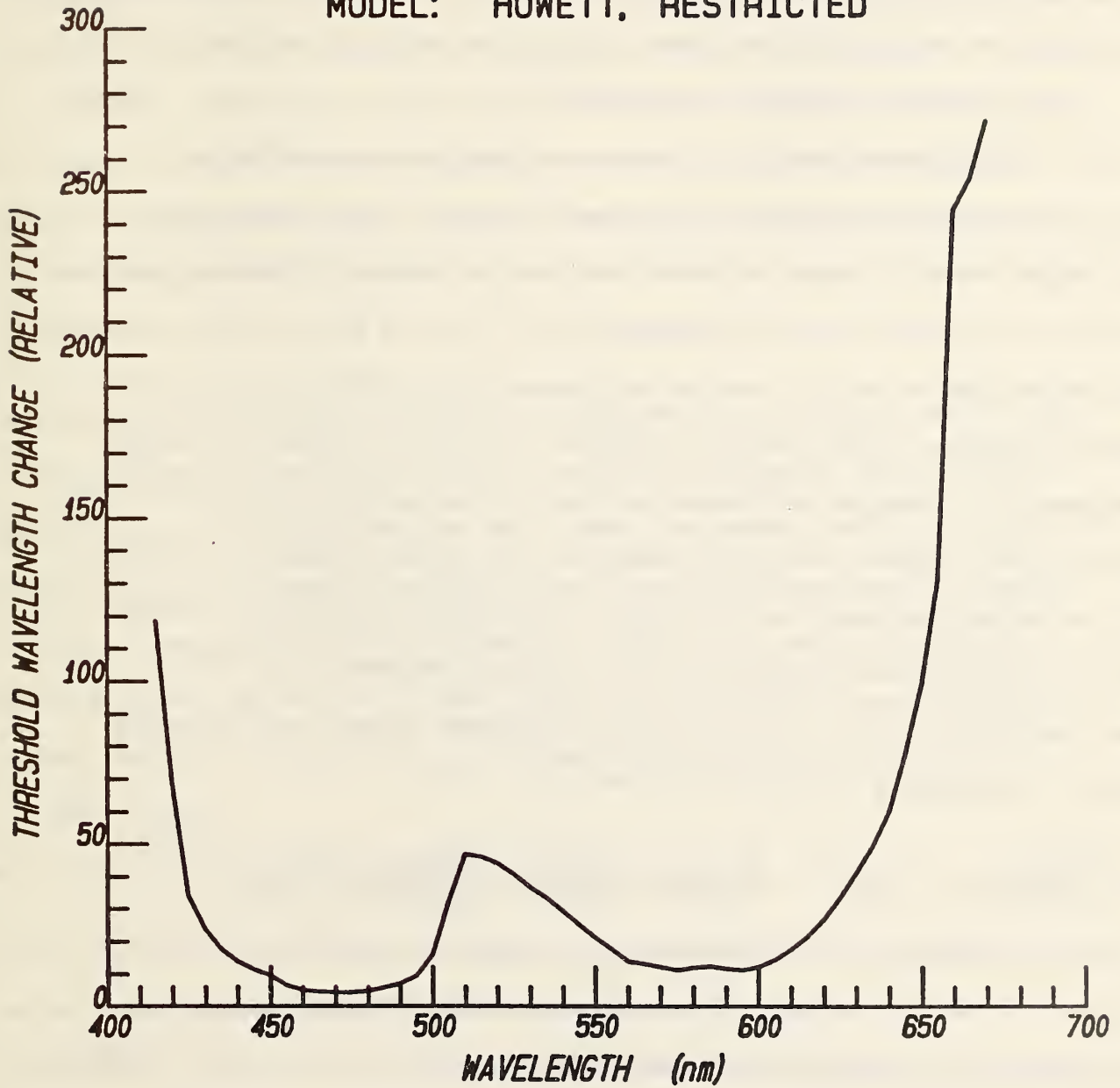


Figure 22. Wavelength discrimination function, as predicted by the restricted Howett model. The units on the ordinate axis are arbitrary.

The mathematical basis for the predicted wavelength-discrimination curves is the assumption, generalized from that made by Guth, that the total psychophysical difference, ΔE , between any two colors A_1, T_1, D_1 and A_2, T_2, D_2 is given by

$$\Delta E = [|A_1 - A_2|^p + |T_1 - T_2|^p + |D_1 - D_2|^p]^{1/p}. \quad (13a)$$

In order to isolate the ability to chromatically discriminate different wavelengths from any brightness differences that might exist between the wavelengths, we deal, again following Guth, with a constant-brightness spectrum. It should be recalled that the functions $\bar{a}(\lambda)$, $\bar{t}(\lambda)$, and $\bar{d}(\lambda)$, of the spectral tristimulus form of Eq. (5), define the opponent-level channel responses for a constant-radiance ("equal energy") spectrum. In order to obtain the channel responses for a constant-brightness spectrum, we impose the condition that the spectrum have constant equivalent luminance. The desired responses to yield unit equivalent luminance are $\bar{a}(\lambda)/V_q(\lambda)$, $\bar{t}(\lambda)/V_q(\lambda)$, and $\bar{d}(\lambda)/V_q(\lambda)$, where $V_q(\lambda)$ is the equivalent luminance of the spectrum, defined in Eq. (13).

Then, applying Eq. (13a), the total color difference between equally bright lights of wavelengths λ and $\lambda + \Delta\lambda$ is given by

$$\Delta E = \left\{ \left| \frac{\bar{a}(\lambda + \Delta\lambda)}{V_q(\lambda + \Delta\lambda)} - \frac{\bar{a}(\lambda)}{V_q(\lambda)} \right|^p + \left| \frac{\bar{t}(\lambda + \Delta\lambda)}{V_q(\lambda + \Delta\lambda)} - \frac{\bar{t}(\lambda)}{V_q(\lambda)} \right|^p + \left| \frac{\bar{d}(\lambda + \Delta\lambda)}{V_q(\lambda + \Delta\lambda)} - \frac{\bar{d}(\lambda)}{V_q(\lambda)} \right|^p \right\}^{1/p}. \quad (13b)$$

For each wavelength, λ , in the spectrum, Eq. (13b) was applied to each model with $\Delta\lambda = +1$ nm and also $\Delta\lambda = -1$ nm; the two ΔE values for each model were averaged; and the average was plotted against λ to produce Figs. 19-22. This is a direct generalization of the procedure specified by Guth and Lodge (1973).

6.6 NEUTRAL POINTS

As can be seen in Figs. 7-10, both the red-green and blue-yellow channels of all opponent-colors models take on a value of zero at one or two wavelengths in the interior (as opposed to the ends) of the visible spectrum. When the

red-green channel is at zero, neither red nor green is being signaled. Hence, the color being perceived must be a pure yellow, a pure blue, or a neutral (white, gray or black) --the only colors that stimulate no perception of either red or green. Similarly, a zero value of the blue-yellow channel implies the perception of a pure red, a pure green, or a neutral. Clearly, when both opponent channels are at zero, the only possible perception is neutral; and, conversely, any neutral (achromatic) color must fail to stimulate both opponent channels.

The mathematical basis for specifying exactly where each opponent channel has the value zero, for the general linear model, is provided by the last two of Eqs. (5): i.e., the equations for the opponent responses T and D as a function of X, Y, and Z. If we set the right side of either of these equations equal to zero, we can translate into chromaticity terms (x, y, and z) by simply dividing through by $X + Y + Z$. When $1-x-y$ is substituted for z, the resulting expressions for the loci in the x,y diagram along which T and D are zero, are respectively

$$T = 0: (a_{21}-a_{23})x + (a_{22}-a_{23})y + a_{23} = 0 \quad (14a)$$

$$D = 0: (a_{31}-a_{33})x + (a_{32}-a_{33})y + a_{33} = 0 \quad (14b)$$

Obviously, both of Eqs. (14) are of the form to represent straight lines in the chromaticity diagram. Moreover, the intersection of the straight lines having the equations (14a) and (14b) is the (unique) point corresponding to a neutral perception.

An explicit formula for the neutral point of the general model can be obtained by solving Eqs. (14) simultaneously, the result being

$$x = (a_{22} a_{33} - a_{23} a_{32})/g \quad (15a)$$

$$y = (a_{23} a_{31} - a_{21} a_{33})/g , \quad (15b)$$

where

$$g = a_{31}(a_{23} - a_{22}) + a_{32}(a_{21} - a_{23}) + a_{33}(a_{22} - a_{21}) . \quad (15c)$$

The only situation in which the neutral point fails to exist is when the two null lines of the opponent channels [Eqs. (14)] are parallel. Algebraically, this condition is equivalent to the quantity g of Eq. (15c) being zero. The non-intersection of the null lines is highly "unnatural," and it seems unlikely that any model incorporating this feature would make color predictions that resemble data adequately.

In the process of deriving Eqs. (14) and (15), we have incidentally provided a geometric proof of an important theorem. We see in Eq. (6) that $\underline{B}$ (equivalent luminance) is equal to $\underline{A}$ (luminance according to the model) if, and only if, both $\underline{T}$ and $\underline{D}$ are zero. We have shown that $\underline{T}$ is zero if and only if the chromaticity point of the color lies on the line represented by Eq. (14a) and $\underline{D}$ is zero if and only if the point lies on the line represented by Eq. (14b).

Hence, $B = A$ --or the ratio $B/A = 1$ -- if and only if the point is the intersection of these two null lines. Since two lines intersect in a single point if they intersect at all, we have established that either (a) the value of B/A is unity at one specific point (the neutral point) and is greater than unity everywhere else [because the two terms other than $\underline{A}$ in Eq. (6) are both non-negative, so that their sum is positive everywhere except at the neutral point];

or, (b) [in the case where the null lines are parallel] there is no point for which $B/A = 1$, and the ratio is greater than unity everywhere.

In representations of experimentally determined constant- B/A (or B/Y or B/L) contours in the literature, there appears to remain some doubt as to whether the locus upon which $B/A=1$ is a single point, a closed curve, or an entire area. However, the preceding arguments show that whatever the actual facts may be, the prediction of the general linear model is that the locus is a single point.

It should be noted that in the preceding derivation, the model has been taken at face value, and the model's own formula for luminance ($\underline{A}$) has been used, rather than the currently defined measure of luminance ($\underline{Y}$). For a model having a definition of $\underline{A}$ not identical to $\underline{Y}$, B/Y will not in general be equal to unity at the neutral point, although B/A must be unity. Such a model is asserting that the CIE 1931 $\bar{y}(\lambda)$ function [= 1924 $V(\lambda)$] is not a perfectly correct representation of the spectral response of the eye's luminance channel. It behooves us to accept all the predictions of a model as is, especially since an error in $\bar{y}(\lambda)$ is widely acknowledged actually to exist (Judd, 1951). Later, in the discussion of the constant- B/A ovals of the general model, it will be seen that there is a compelling computational simplicity associated with the use of $\underline{A}$ rather than $\underline{Y}$ for luminance (when the two are not the same).

6.6.1 Guth and Lodge

Figure 23 shows that in the Guth and Lodge model, the line (dashed) of no redness-greenness extends from approximately the shortwave end of the spectrum locus to a region that can reasonably be described as yellow. As explained in

MODEL: GUTH & LODGE

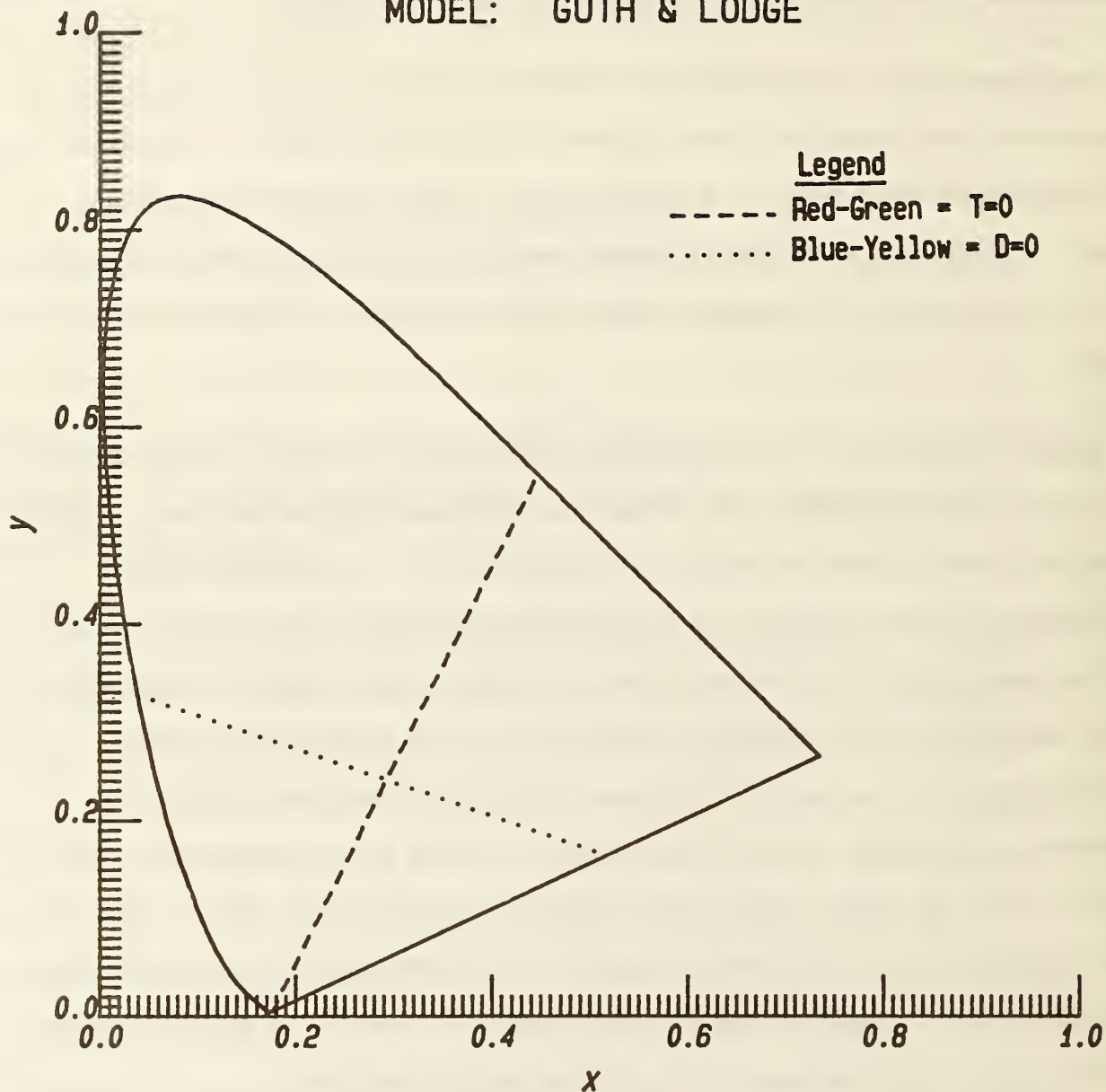


Figure 23. Loci of zero red-green (dashed) and zero blue-yellow (dotted), as implied by the Guth and Lodge model. The intersection of the two lines is the white (neutral, achromatic) point of the model.

the opening paragraph of section 6.6, this line must contain the chromaticity points of colors perceived as pure blue, white, and pure yellow. Unfortunately, the shortwave end of the spectrum is not blue but violet; i.e., the perceived color contains a definite admixture of redness. Thus, a well-known defect of the Guth model is pointed up by the figure.

The line (dotted) of no blueness-yellowness has both ends in not quite the right places. The stimulus producing a perception of pure red does indeed lie on the purple boundary, rather than at the longwave end of the spectrum. The point where the dotted line in Fig. 23 hits the purple boundary is, however, too far toward the violet end, and corresponds to a purplish red with a definite component of perceived blue. Similarly, the other end of the dotted line cannot be described as the pure green wavelength of the spectrum; it corresponds to a clearly bluish green.

Finally, the neutral point of the Guth and Lodge model --the intersection of the dashed and dotted lines-- is also a little on the bluish side, perceptually. In the later version of the model (Guth, Massof and Benzsshawel, 1980), the definitions of the opponent channels were modified in a way that resulted in a neutral point closer to what is usually considered pure white. As explained earlier, the present author has not attempted to explore the properties of the latter model, since the choice of adaptation constants changes most of the implications of the model, including the neutral point.

6.6.2 Thornton

Figure 24 reveals a serious defect of the Thornton model. The line (dashed) of no redness-greenness extends from the blue-green part of the spectrum to the

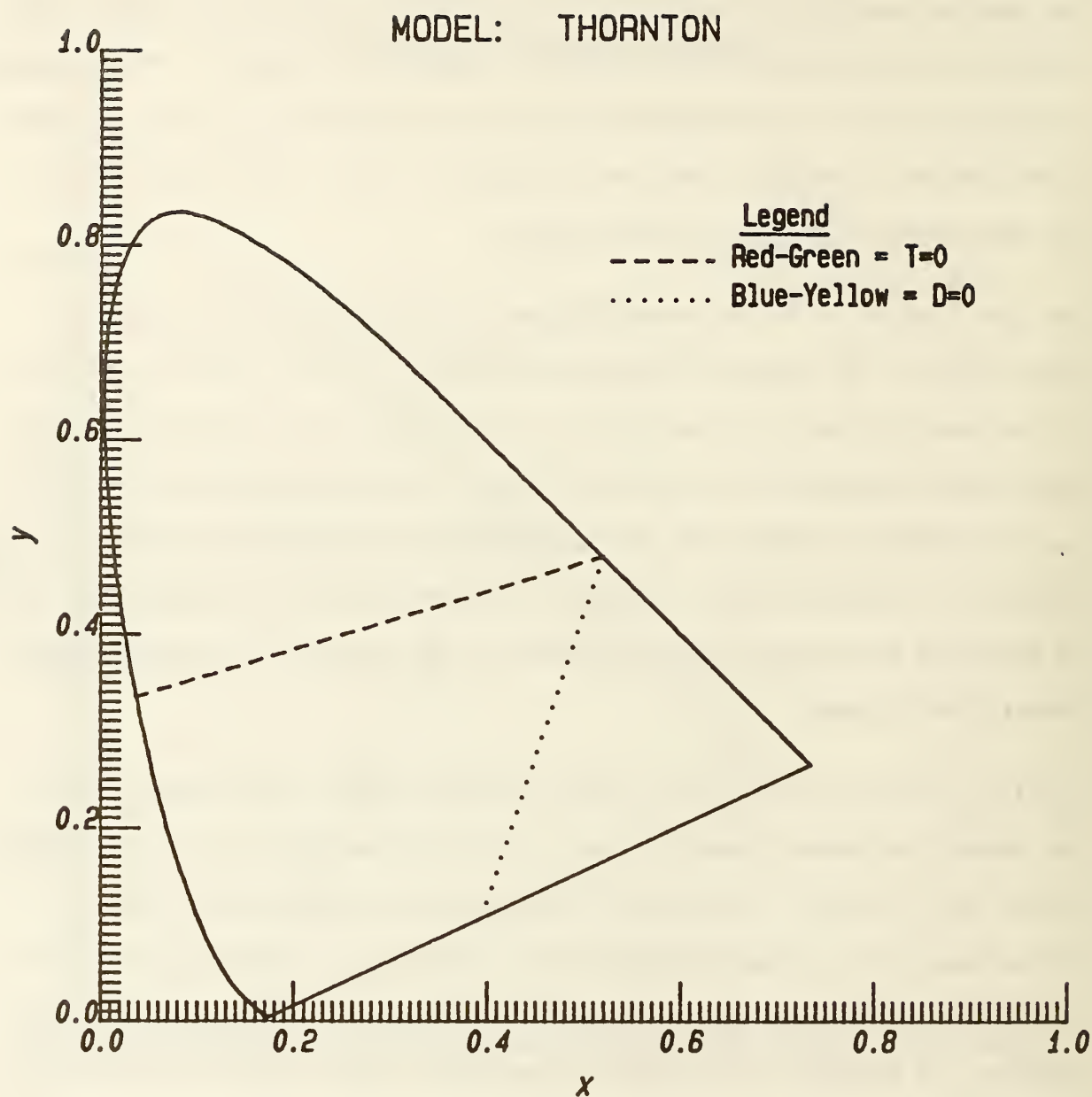


Figure 24. Loci of zero red-green (dashed) and zero blue-yellow (dotted), as implied by the Thornton model. The intersection of the two lines is the white (neutral, achromatic) point of the model, but in this case the point has the chromaticity of a spectrally pure yellow.

yellow; and the line (dotted) of no blueness-yellowness extends from the purple to the yellow. The result is that the "neutral" point of the Thornton model, instead of being in the central region perceived as white, is actually a spectrally pure yellow, of wavelength very close to 581 nm. (Careful computation shows that the point corresponds to an imaginary color just beyond the spectrum locus, but negligibly so.)

6.6.3 Howett, Best-Fitting

The dashed line in Fig. 25, which should join the pure blue part of the spectrum locus to the pure yellow part, does fall in approximately the right location. The dotted line, which should join the pure red and pure green points, is clearly in the wrong place. It joins orangish red to a slightly bluish green. Considering that this model was derived entirely on the basis of brightness matches, its performance on this test is by no means disgraceful, compared to that of the Guth and Lodge or Thornton models.

6.6.4 Howett, Restricted

Comparison of Figs. 25 and 26 shows that the author's model restricted to give an all-positive luminance-channel function is a bit worse than the very best-fitting model with regard to the opponent-channel null loci. Both lines are rotated slightly in the wrong direction. That is, in the restricted model, the blue point is further toward violet, and the green point is further toward blue. The red point --the right end of the dotted line-- is still a quite orangish red. The neutral points of both models are within the region usually identified as white, toward the yellowish side, with the restricted model's white point being a bit less yellow.

MODEL: HOWETT, BEST-FITTING

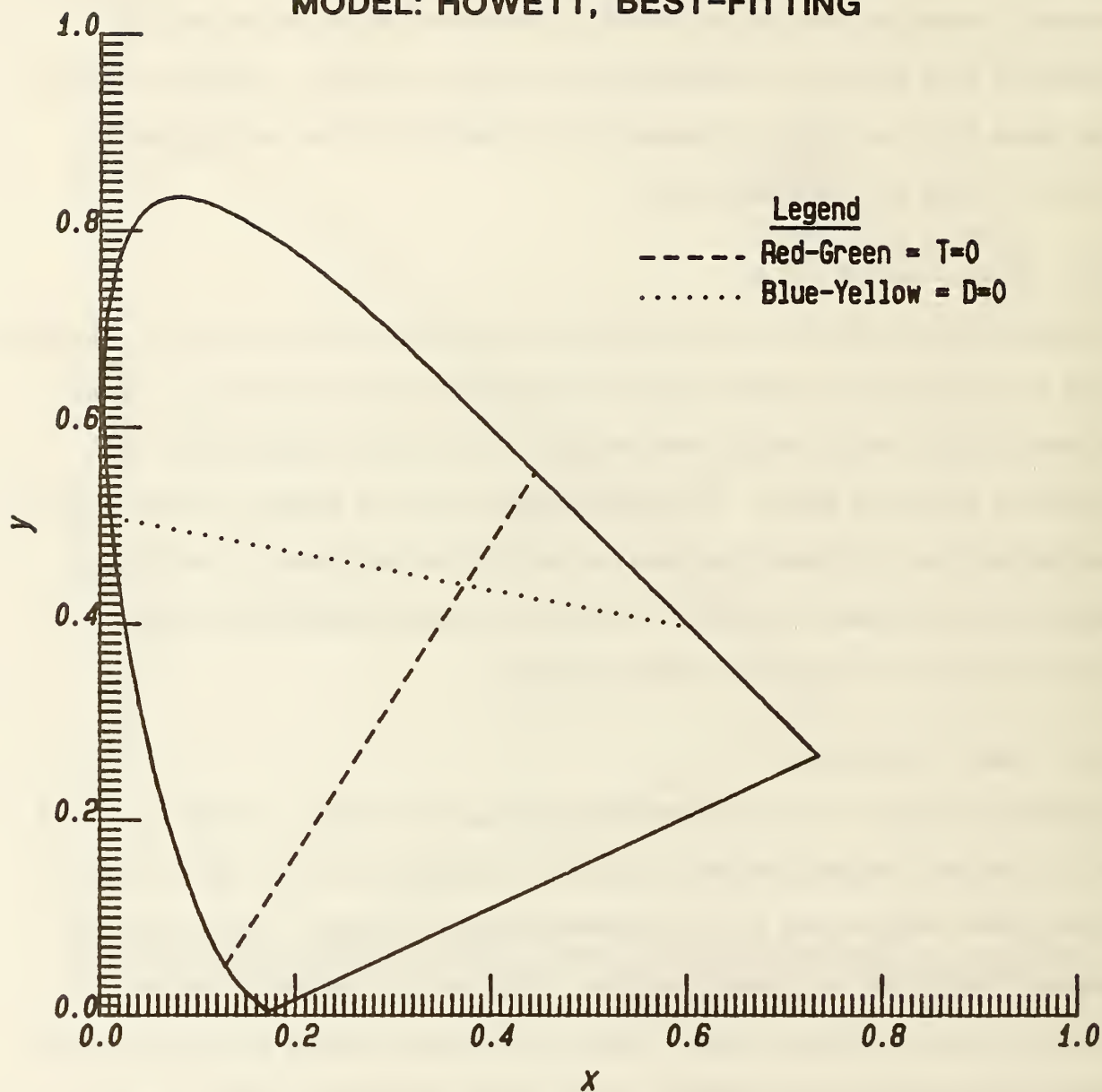


Figure 25. Loci of zero red-green (dashed) and zero blue-yellow (dotted), as implied by the best-fitting Howett model. The intersection of the two lines is the white (neutral, achromatic) point of the model.

MODEL: HOWETT, RESTRICTED

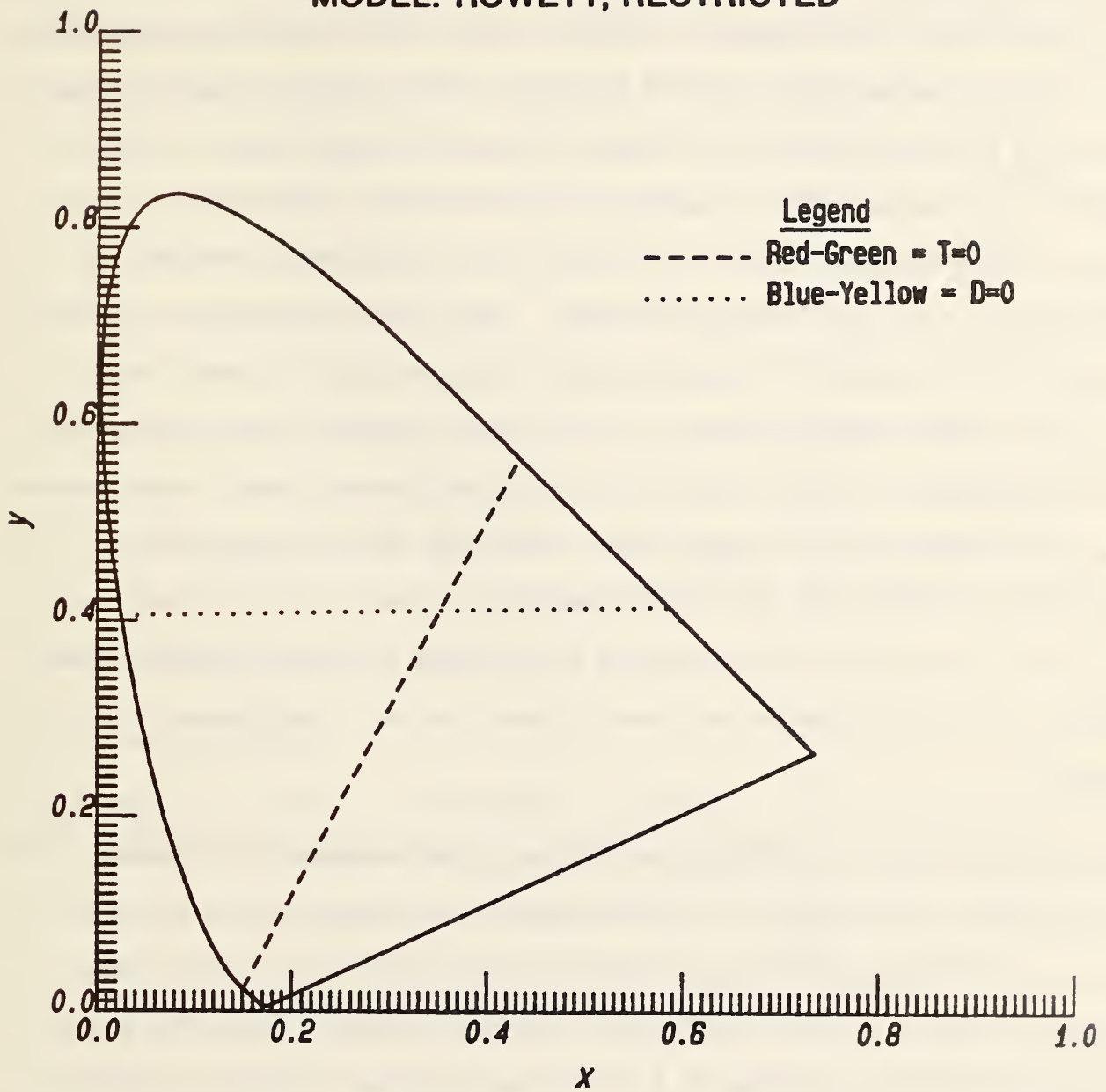


Figure 26. Loci of zero red-green (dashed) and zero blue-yellow (dotted), as implied by the restricted Howett model. The intersection of the two lines is the white (neutral, achromatic) point of the model.

6.7 THE PRIMARIES

In a traditional, three-component (receptor-stage) color system, without opponent channels, the chromaticity of one of the three primary colors of the system can be located by setting both of the other two primaries equal to zero. (This leads to pure "stimulation" by only the third primary). This procedure is precisely what was done [Eqs. (14)] in the previous subsection to locate the neutral point of an opponent-colors system. Two of the three channel responses (namely the two chromatic, opponent channels) were set equal to zero. It is natural to wonder about the meaning of the points located by setting equal to zero both members of each of the two other possible channel pairs. Thus, colors that can be identified by analogy as the "primaries" of an opponent-colors model can be derived from the three conditions $T = D = 0$, $A = T = 0$, and $A = D = 0$. The author is not currently able to give an interpretation of what meaning is to be attached to the three "primary" colors of an opponent-colors system.

As we saw with Eqs. (14), setting two of the channel responses of a linear opponent-colors model equal to zero corresponds to intersecting two straight lines on the chromaticity diagram, thereby determining a unique point. The meaning of a zero value for one of the opponent (chromatic) channels is clear, but what is the interpretation of $A = 0$? Since A is the luminance channel, the line $A = 0$ is the locus of colors that the model assigns zero luminosity; in short, it is the "lightless line" or alychne of traditional, three-component chromaticity diagrams. The alychne of the CIE 1931 x,y diagram is the line $y = 0$, which of course lies wholly outside of the region of real colors; all real colors have positive luminosity. Since every point on the line $A = 0$

therefore corresponds to an imaginary color, we know that the intersection of any other line with $A = 0$ determines a chromaticity point corresponding to an imaginary color. In other words, we know in advance that two of the three "primaries" of an opponent-colors system are always imaginary. The third "primary," determined by intersecting the lines $T = 0$ and $D = 0$, can be either real or imaginary, depending on the definitions of the opponent channels associated with the particular model. Ordinarily, since the point corresponding to the conditions $T = D = 0$ is the neutral point of the model (section 6.6), it should be expected that the third primary --the one corresponding to pure stimulation of only the A channel--should be real and should correspond to a white color. This reasoning may help to make clearer why the A channel is spoken of in different contexts as being alternatively the luminance or the white (achromatic, neutral) channel.

6.7.1 Guth and Lodge

The alychne of the Guth and Lodge model is close to $y = 0$, as seen in Fig. 27. This reflects the fact that Guth's A channel is defined [Eq. (1)] as very close to Y (or $\bar{y}$). Any reasonable model that explicitly identified its A channel as corresponding to luminance would be expected to define A as nearly or actually equal to Y, since $\bar{y}$ is known to be close to the correct weighting function for calculating luminance. This will be confirmed graphically in the other figures of this subsection, as well as in the accompanying tables.

It is important to recall that in deriving the author's best-fitting model, no assumptions at all were made about the nature of the three channels, beyond that there were three channels and that the spectral response of each of them was expressible as a linear combination of $\bar{x}$, $\bar{y}$, and $\bar{z}$. Thus, it is significant

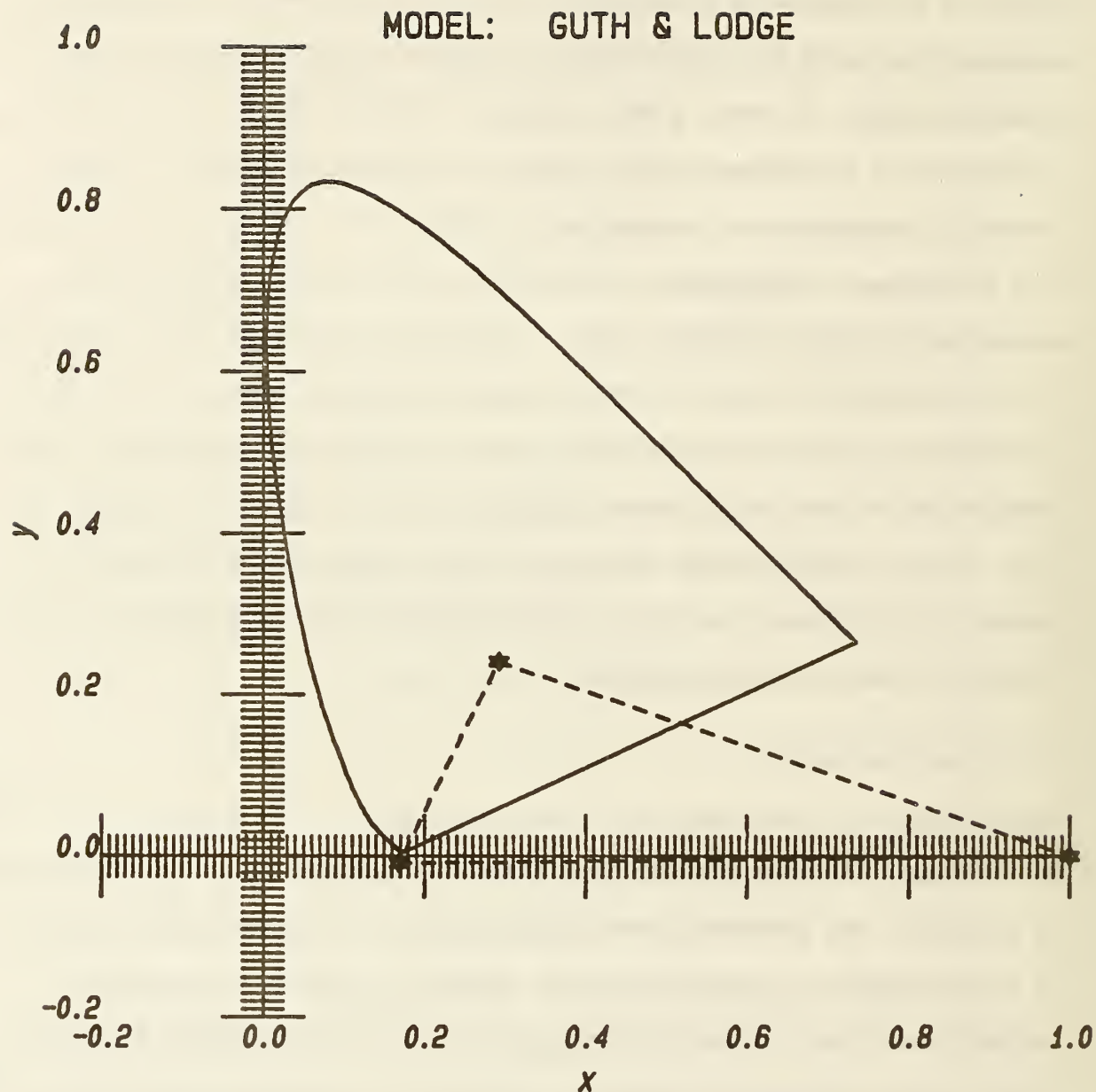


Figure 27. The opponent-level primaries (stars) of the Guth and Lodge model. These are the points for which two of the channel responses A , T , and D are zero. The upper two lines of the triangle ($T=0$ and $D=0$) are the same as those in Fig. 23; the bottom line is $A=0$ (the alychne of the model).

that what emerged from the optimization was one almost entirely positive channel whose response clearly resembled $\bar{y}$; and two opponent channels that made reasonable sense as red-green and blue-yellow responses. This outcome suggests that accounting for brightness-matching data may require a model that considerably resembles contemporary opponent-color theories.

Even the author's restricted model was subject only to the single requirement that one of the channels have all-positive response. There was no further requirement that this channel have a spectral response at all resembling $\bar{y}$, in the sense that any positive coefficients of $\bar{x}$ and $\bar{z}$ could have emerged. Nevertheless, the data forced the model to have its all-positive channel essentially identical to $\bar{y}$ [see Eqs. (9)], and again the other two channels took on opponent forms that could not unreasonably be identified as red-green and (perhaps somewhat less clearly) blue-yellow. Again, this result can be construed as suggesting the possible reality of the basic opponent-colors notion.

Moreover, it should not be forgotten that in Guth's earliest work, which involved the factor analysis of a matrix of data representing the additivities at threshold of lights of different wavelengths, he too found that the data implied the existence of a channel with a response resembling $\bar{y}$, and two opponent channels. Thus, the opponent-colors model, in broad outline, seems to be more than an abstraction that can be made to account for certain vision data; instead, more than one kind of vision data seems to directly imply the essential truth of that model.

Returning now to the specific characteristics of the Guth and Lodge model, we observe that the point (indicated by a star) at the lower left corner of the

triangle of primaries in Fig. 27 is the intersection of the lines corresponding to $A = 0$ and (see Fig. 23) $T = 0$. The point therefore determines the primary corresponding to pure stimulation of the D or blue-yellow channel. This point is just outside the shortwave end of the spectrum locus, in the same area where the tritanopic convergence (sometimes called "co-punctal") point (the missing blue primary of the tritanope) is traditionally located. The lower right point of the primary triangle, which corresponds to the red-green primary, is located very nearly on an extension of the longwave end of the spectrum locus --the line $z = 0$ or $x + y = 1$. It is in a region where some authors locate the convergence (co-punctal) point (missing primary) of deuteranopes. Its color would therefore be ordinarily identified as green (not red).

Table 1 provides accurate numerical information, for the Guth and Lodge model, on both the equations of the null lines and the coordinates of the primaries. One primary of any of these models is always (at least nominally) white, but it is not obvious what criterion should be used to decide whether an imaginary red-green primary of an opponent-colors system should be labeled either red or green, particularly when it is located directly on the alychne of the system. Consequently, the practice followed in Table 1, and in the analogous tables for the other three models, was to use the non-committal terms "red-green primary" and "blue-yellow primary".

6.7.2 Thornton

The primaries of Thornton's model, as revealed in Fig. 28, are quite different from those of the Guth and Lodge model (Fig. 27). What is presumably supposed to be a blue primary (lower left) is far outside the spectrum locus, and is in

Table 1. Guth and Lodge Model: Null Lines and Primaries of Opponent-Level Channels

$$A = 0: y = 0.010593x - 0.010593$$

Wavelength^a x y
Outside region of real colors.

$$T = 0: y = 2.016702x - 0.348643$$

<u>Wavelength^a</u>	<u>x</u>	<u>y</u>
360.739	0.17550	0.00529
570.371	0.44666	0.55213

$$D = 0: y = -0.340909x + 0.340909$$

<u>Wavelength^a</u>	<u>x</u>	<u>y</u>
491.470	0.03862	0.32774
C491.470	0.51773	0.16441

<u>Primary</u>	<u>Conditions</u>	<u>x</u>	<u>y</u>
Blue-yellow	A = T = 0	0.168510	-0.008808
Red-green	A = D = 0	1.000000	0.000000
White	T = D = 0	0.292479	0.241200

^a The wavelengths (nm) and corresponding chromaticities are the intersection points of the null lines with the spectrum locus or purple boundary. The letter C denotes "complementary" with respect to the neutral (white) point of this model, and indicates a point on the purple line.

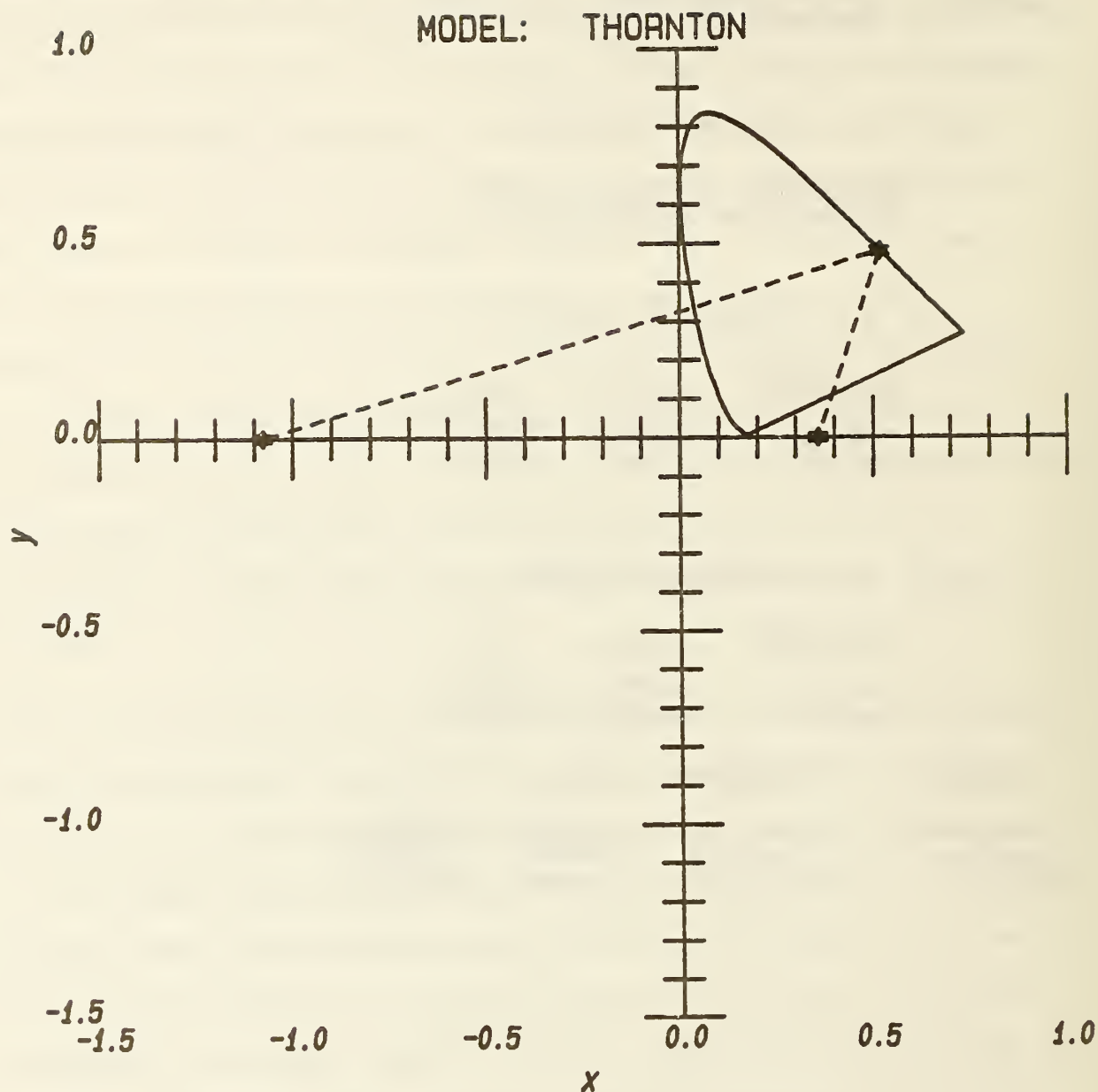


Figure 28. The opponent-level primaries (stars) of the Thornton model. These are the points for which two of the channel responses A, T, and D are zero. The upper two lines of the triangle ($T=0$ and $D=0$) are the same as those in Fig. 24; the bottom line--here coinciding with the x-axis-- is $A=0$ (the alychne of the model).

more of a blue-green direction. Neither this nor the other chromatic primary (lower right) is in a location that is ordinarily thought of as a convergence (co-punctal) point for any type of dichromatic vision. The direction of the red-green primary is here actually toward the purple. As has been previously noted (in 6.6.2), the "white" primary is in fact a spectral yellow in Thornton's model.

The numerical specifications of Thornton's null lines and primaries are given in Table 2. The null lines $T = 0$ and $D = 0$ have very nearly --but not quite-- converged in the neighborhood of 581 nm on the spectrum locus; the intersection point is barely beyond the locus.

6.7.3 Howett, Best-Fitting

In Fig. 29, the red-green primary is very far off to the right. More disturbingly, the blue primary (lower left) is a real color, in contradiction of the assertion made in the introductory remarks of subsection 6.7. The dashed line just above the x-axis ($y=0$) is the alychne, yet it intersects a corner area of the region of real colors! The consequence is that some real colors (violets) are assigned zero and negative luminosities (the latter being the colors within the "horseshoe" that lie below the alychne). This physically impossible situation is a consequence of the negative lobe of the A function for this model (Fig. 9), and is the reason the model was cast aside.

Table 3 summarizes the numerical specifications of the null lines and primaries for this model.

Table 2. Thornton Model: Null Lines and Primaries of Opponent-Level Channels

$$A = 0: y = 0$$

<u>Wavelength^a</u>	<u>x</u>	<u>y</u>
Outside region of real colors.		

$$T = 0: y = 0.301000x + 0.323805$$

<u>Wavelength^a</u>	<u>x</u>	<u>y</u>
491.786	0.03718	0.33499
580.998	0.51906	0.48004

$$D = 0: y = 2.976110x - 1.064979$$

<u>Wavelength^a</u>	<u>x</u>	<u>y</u>
581.008	0.51912	0.47998
C581.008	0.39371	0.10674

<u>Primary</u>	<u>Conditions</u>	<u>x</u>	<u>y</u>
Blue-yellow	A = T = 0	-1.075765	0.000000
Red-green	A = D = 0	0.357842	0.000000
White	T = D = 0	0.519150	0.480069

^a The wavelengths (nm) and corresponding chromaticities are the intersection points of the null lines with the spectrum locus or purple boundary. The letter C denotes "complementary" with respect to the neutral (white) point of this model, and indicates a point on the purple line.

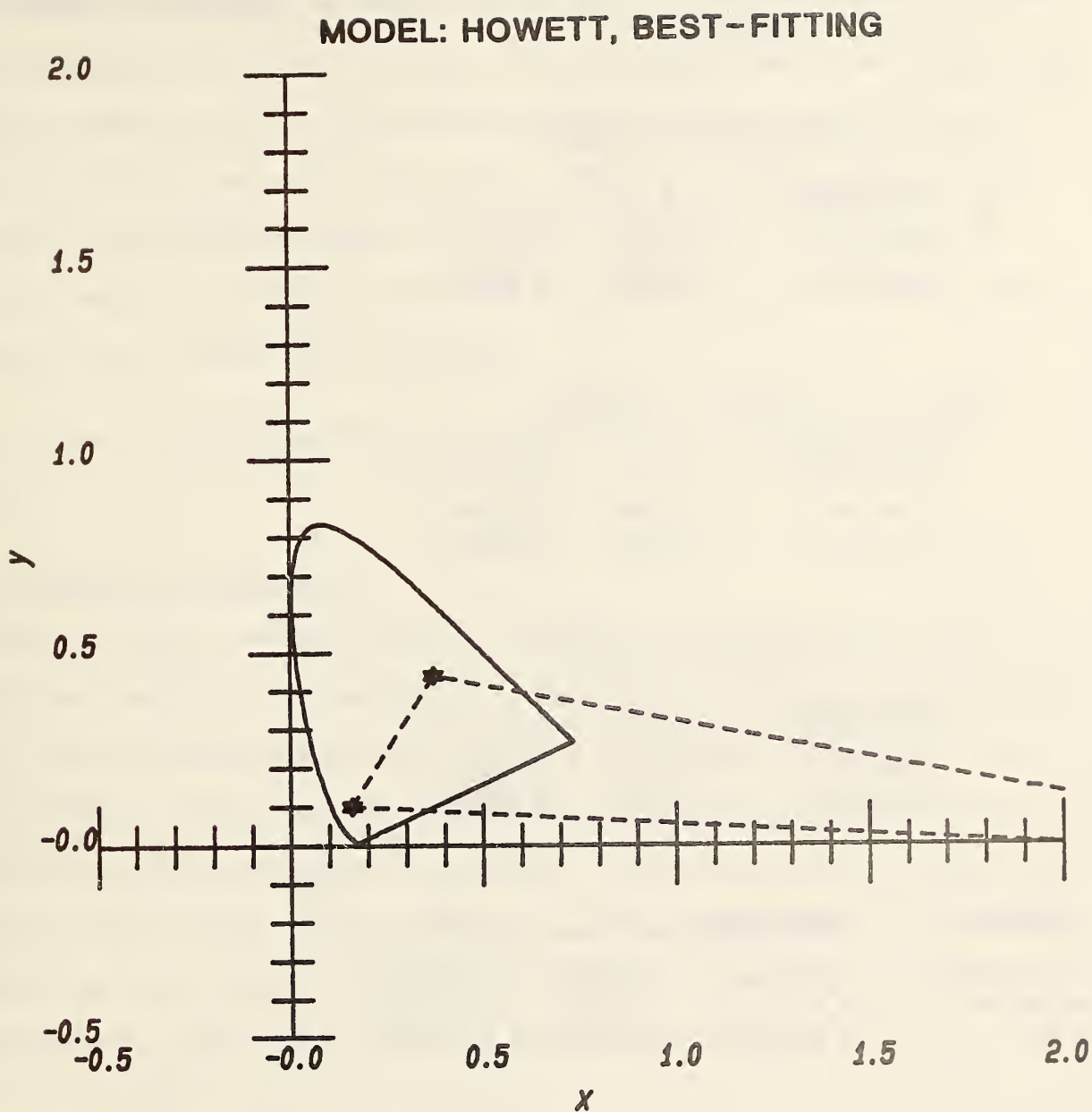


Figure 29. The opponent-level primaries (stars) of the best-fitting Howett model. One point (the red-green primary) is off the figure to the right. These are the points for which two of the channel responses A, T, and D are zero. The upper two lines of the triangle ($T=0$ and $D=0$) are the same as those in Fig. 25; the bottom line is $A=0$ (the alychne of the model).

Table 3. Howett Best-Fitting Model: Null Lines and Primaries of Opponent-Level Channels

$$A = 0: y = -0.054475x + 0.108950$$

<u>Wavelength^a</u>	<u>x</u>	<u>y</u>
477.055	0.10258	0.10336
C477.055	0.35668	0.08952

$$T = 0: y = 1.589946x - 0.153770$$

<u>Wavelength^a</u>	<u>x</u>	<u>y</u>
468.258	0.12846	0.05048
570.135	0.44501	0.55377

$$D = 0: y = -0.190283x + 0.510433$$

<u>Wavelength^a</u>	<u>x</u>	<u>y</u>
498.799	0.01102	0.50834
595.193	0.60392	0.39552

<u>Primary</u>	<u>Conditions</u>	<u>x</u>	<u>y</u>
Blue-yellow	A = T = 0	0.159765	0.100247
Red-green	A = D = 0	2.956245	-0.052092
White	T = D = 0	0.373100	0.439438

^a The wavelengths (nm) and corresponding chromaticities are the intersection points of the null lines with the spectrum locus or purple boundary. The letter C denotes "complementary" with respect to the neutral (white) point of this model, and indicates a point on the purple line.

6.7.4 Howett, Restricted

The blue primary in Fig. 30 is in about the same position as it has in the Guth and Lodge model (Fig. 27). The white primary is reasonably situated, in a slightly yellowish part of the white area. The most striking feature is the location of the red-green primary very far off to the left. The author is currently unable to interpret this result or correlate it with any particular features of the predictions of this model.

Table 4 gives the numerical data for the null lines and primaries of this model.

6.8 CONTOURS OF CONSTANT B/A

In the past, authors who have wanted to exhibit contours of constant B/A (equivalent luminance over luminance) on the chromaticity diagram, for their own model, have calculated B/A at each point of a fairly finely spaced grid, and then sketched in each contour by interpolating smoothly by eye along a path passing near points with similar B/A values. This is the same method used with empirical data, such as elevation contours on topographical maps, or isobars and isotherms on weather maps. However, it is possible to derive an explicit equation for the locus $B/A = k$ ($k > 1$), for the general linear model.

Applying Eqs. (5) and (6), we can write

$$B^P = |a_{11}X + a_{12}Y + a_{13}Z|^P + |a_{21}X + a_{22}Y + a_{23}Z|^P + |a_{31}X + a_{32}Y + a_{33}Z|^P. \quad (16)$$

We want to set B/A equal to the specific value $\underline{k}$; i.e.,

$$B = kA, \quad (17)$$

or

$$B^P = k^P A^P. \quad (18)$$

1.0



-2.0

Table 4. Howett Restricted Model: Null Lines and Primaries of Opponent-Level Channels

$$A = 0: y = -0.001521x$$

Wavelength^a x y
Outside region of real colors.

$$T = 0: y = 1.921231x - 0.265231$$

<u>Wavelength^a</u>	<u>x</u>	<u>y</u>
455.571	0.15025	0.02343
568.379	0.43266	0.56601

$$D = 0: y = 0.016667x + 0.404167$$

<u>Wavelength^a</u>	<u>x</u>	<u>y</u>
494.672	0.02478	0.40458
591.795	0.58548	0.41392

<u>Primary</u>	<u>Conditions</u>	<u>x</u>	<u>y</u>
Blue-yellow	A = T = 0	0.137943	-0.000210
Red-green	A = D = 0	-22.222600	0.033790
White	T = D = 0	0.351470	0.410025

^a The wavelengths (nm) and corresponding chromaticities are the intersection points of the null lines with the spectrum locus or purple boundary. The letter C denotes "complementary" with respect to the neutral (white) point of this model, and indicates a point on the purple line.

The right sides of Eqs. (16) and (18) are both equal to B^P , and hence are equal to each other:

$$k^{PA^P} = |a_{11}X + a_{12}Y + a_{13}Z|^P + |a_{21}X + a_{22}Y + a_{23}Z|^P + |a_{31}X + a_{32}Y + a_{33}Z|^P. \quad (19)$$

Since we know from Eq. (5) that

$$A = |a_{11}X + a_{12}Y + a_{13}Z|, \quad (20)$$

and therefore that

$$A^P = |a_{11}X + a_{12}Y + a_{13}Z|^P, \quad (21)$$

we can replace A^P in Eq. (19) by its equivalent given by Eq. (21). The result is

$$k^P |a_{11}X + a_{12}Y + a_{13}Z|^P = |a_{11}X + a_{12}Y + a_{13}Z|^P + |a_{21}X + a_{22}Y + a_{23}Z|^P + |a_{31}X + a_{32}Y + a_{33}Z|^P, \quad (22)$$

or

$$(k^P - 1) |a_{11}X + a_{12}Y + a_{13}Z|^P = |a_{21}X + a_{22}Y + a_{23}Z|^P + |a_{31}X + a_{32}Y + a_{33}Z|^P. \quad (23)$$

Equation (23) can be converted from tristimulus values to chromaticity coordinates by dividing the entire equation through by $|X + Y + Z|^P$. The result is

$$(k^P - 1) |a_{11}x + a_{12}y + a_{13}z|^P = |a_{21}x + a_{22}y + a_{23}z|^P + |a_{31}x + a_{32}y + a_{33}z|^P. \quad (24)$$

Finally, substituting for $\underline{z}$ its equivalent,

$$z = 1 - x - y, \quad (25)$$

and rearranging terms, we arrive at the final, explicit equation

$$(k^P - 1) |(a_{11} - a_{13})x + (a_{12} - a_{13})y + a_{13}|^P = |(a_{21} - a_{23})x + (a_{22} - a_{23})y + a_{23}|^P + |(a_{31} - a_{33})x + (a_{32} - a_{33})y + a_{33}|^P. \quad (26)$$

Because of the power, p , as well as the absolute values, Eq. (26) cannot be explicitly solved for $\underline{y}$ as a function of $\underline{x}$.

6.8.1 The Special Case $p=2$ (Including the Guth Model)

In order to use Eq. (26) to generate the $B/A=k$ contours for a particular model,

it would appear that some sort of iterative numerical procedure must be employed in order to obtain the value of y corresponding to a given x . However, there is one special case in which Eq. (26) reduces to a recognizable form: namely, $p=2$, as in the models of Guth and of Ingling. Squaring an absolute value permits the dropping of the absolute value signs, so in this case Eq. (26) reduces to an ordinary second degree equation in x and y . Such an equation represents a conic section. Because the B/A ovals are closed curves in the neighborhood of the white point of the model ($B/A = 1$), we know that Eq. (26) must represent an ellipse in that region. It is conceivable that at some distance from the white point (i.e., for large k values), the discriminant of Eq. (26) changes sign, and the locus becomes a different category of conic, such as a hyperbola or parabola, but this possibility has not been examined algebraically by the author.

6.8.2 The Special Case $p=1$ (Including the Thornton Model)

One other case, $p=1$, is slightly less obvious, but also leads to loci of an easily specified form. If we could disregard the absolute value signs in Eq. (26), with $p=1$, the result would be a collection of terms of three types: a constant times x ; a constant times y ; and simply a constant. Combining these terms accordingly leads to the equation of a straight line in the chromaticity diagram.

However, we are not free to disregard the absolute value signs in Eq. (26). Each of the two expressions within the absolute value signs on the right side can be either positive or negative (or zero) for real colors. Fortunately, this leads to only four distinct cases: the signs of these expressions can be

++, +-, -+, or --. No other possibilities exist. (The zero case is not significant, and can be lumped with either + or - or both.) By assigning explicit signs to the expressions inside, we can then drop the absolute values, because we know that for any quantity s , $|s|=s$ if $s \geq 0$; and $|s|=-s$ if $s \leq 0$.

The term inside the absolute value signs on the left side of Eq. (26) is always positive for real colors, for any model in which the alychne does not intersect the region of real colors; which is to say, for any reasonable model. Hence, those absolute value signs can also be dropped. The result, when all the terms are combined, is, in each of the four cases, the equation of a straight line. Points lying on any one of these four lines satisfy Eq. (26). Thus, we have established that for $p=1$, the loci $B/A=k$ are formed from four straight lines; i.e., they are quadrilaterals rather than smoothly curved ovals.

If a model in which some real colors are permitted to have negative luminosity is to be considered, then two more possibilities are introduced: the expression within the absolute value signs on the left side of Eq. (26) can also be either positive or negative. Combining this dichotomy with the four sign possibilities arising on the right side of the equation leads to a total of eight cases ($2 \times 2 \times 2$), each of which leads to the equation of a different straight line. In this situation, the loci of constant B/A are octagons, or possibly some more complex eight-sided figures, such as separated pairs of quadrilaterals.

6.8.3 The General Case (Including the Howett Model)

So far, except when $p=1$, it does not appear as if Eq. (26) can be used easily to trace out constant- B/A contours with precision. However, it turns out, perhaps surprisingly, that Eq. (26) can be converted into a form that easily permits direct, point-by-point plotting through an explicit formula, without

resort to iteration. The unique status of the white point (primary) of a model--the sole point for which $B/A=1$ --suggests that setting up a polar coordinate system centered there might lead to some kind of simplification. Indeed, converting to a polar form allows the separation of variables, and the derivation of an explicit formula for the radius as a function of the angle. The formula can be written more compactly by introducing abbreviations for some of the key quantities in Eq. (26). Let

$$R_p = (k^p - 1)^{1/p} \quad (27)$$

$$\delta_{11} = a_{11} - a_{13}, \quad \delta_{12} = a_{12} - a_{13} \quad (28)$$

$$\delta_{21} = a_{21} - a_{23}, \quad \delta_{22} = a_{22} - a_{23} \quad (29)$$

$$\delta_{31} = a_{31} - a_{33}, \quad \delta_{32} = a_{32} - a_{33}. \quad (30)$$

Then the formula is

$$r' = \frac{R_p(\delta_{11}x_0 + \delta_{12}y_0 + a_{13})}{\{|\delta_{21}\cos\theta' + \delta_{22}\sin\theta'|^p + |\delta_{31}\cos\theta' + \delta_{32}\sin\theta'|^p\}^{1/p} - R_p(\delta_{11}\cos\theta' + \delta_{12}\sin\theta')}, \quad (31)$$

where x_0, y_0 are the chromaticity coordinates of the white point of the model; and r', θ' are polar coordinates in the shifted system centered at x_0, y_0 .

Equation (31) follows from Eq. (26) by applying to Eq. (26) the inverse translation equations

$$x = x' + x_0, \quad y = y' + y_0; \quad (32)$$

and then converting to polar coordinates relative to the new origin, in accordance with the standard equations

$$x' = r'\cos\theta', \quad y' = r'\sin\theta'. \quad (33)$$

The rest of the derivation is algebraic manipulation. Account is taken of the fact that the white point lies on both of the lines $T=0$ and $D=0$, and hence

x_0, y_0 satisfies both Eq. (14a) and Eq. (14b). (This reduces several terms to zero.) The final complete separability of r' seems to be valid for any color for which $\underline{A}$, the luminance of the model, is positive; i.e., for any real color, in a reasonable model. The author has not explored algebraically the validity of Eq. (31) for imaginary colors with negative luminosity.

Equation (31) has been used to generate the contours in Figs. 31-34, for the four models being investigated. (Being entirely general, it applies to the cases $p=1$ and $p=2$, without any special considerations, as well as to all other p values.) The primed coordinates of Eq. (31) can be converted back into unprimed CIE chromaticity coordinates for plotting, by applying Eq. (33) to r', θ' and then applying Eq. (32) to x', y' .

Although Eq. (31) is a valid formula, certain rather unexpected difficulties arise in its actual application. The source of these difficulties is not yet understood. In essence, false sections of some of the curves are generated, a phenomenon somewhat analogous to the appearance of physically meaningless roots of some polynomial equations. These "ghost" curve segments tend to appear above or below the region of real colors, but sometimes encroach upon that region.

The problem appears to be associated with the fractional form of Eq. (31), as opposed to the original, non-fractional Eq. (26). The denominator of Eq. (31) can be positive, negative, or zero. When it is near zero, Eq. (31) calls for an exceedingly large radius, and trouble seems to stem at least partly from this source.

There is a relatively simple solution to the problem. In plotting the oval

$B/A=k$, examine each point generated by formula (31). Using the numerical versions of Eqs. (5) and (6) for the particular model, the actual value of B/A can be calculated for the proposed point. If the radius r' is negative, or if the actual value of B/A for the proposed point differs excessively from the value k supposedly characterizing all points on the curve, reject the point and simply plot no point for that value of θ' . In drawing Figs. 31-34 (by computer), any point was rejected if its actual B/A ratio differed from the nominal value by more than 5×10^{-5} (in either direction).

6.8.4 Guth and Lodge

As anticipated in section 6.8.1, this model, with $p=2$, yields smooth curves which are conic sections (Fig. 31). The most striking feature of Fig. 31 is the seemingly rather abrupt transition, somewhere around a B/A value of 1.0035, from closed ellipses near the white point, to what appear to be hyperbolas, with very straight asymptotic tails. The possibility of such a shift was pointed out in section 6.8.1, but the values of the appropriate constants needed to confirm this effect have not been calculated, either for the general linear model with $p=2$, or for the Guth and Lodge model in particular.

6.8.5 Thornton

As anticipated in section 6.8.2, the Thornton model, with $p=1$, yields constant- B/A "curves" which are straight-sided quadrilaterals (Fig. 32). Thornton himself (1973) was aware of this implication of his model, but chose to grossly round off the corners in presenting the corresponding figure in his 1973 paper. Presumably, this was because he believed that nature tends not to have sharp corners, except at the quantum level. There is a very slight rounding of some of the corners in Fig. 32, but this effect is an artifact of the computer

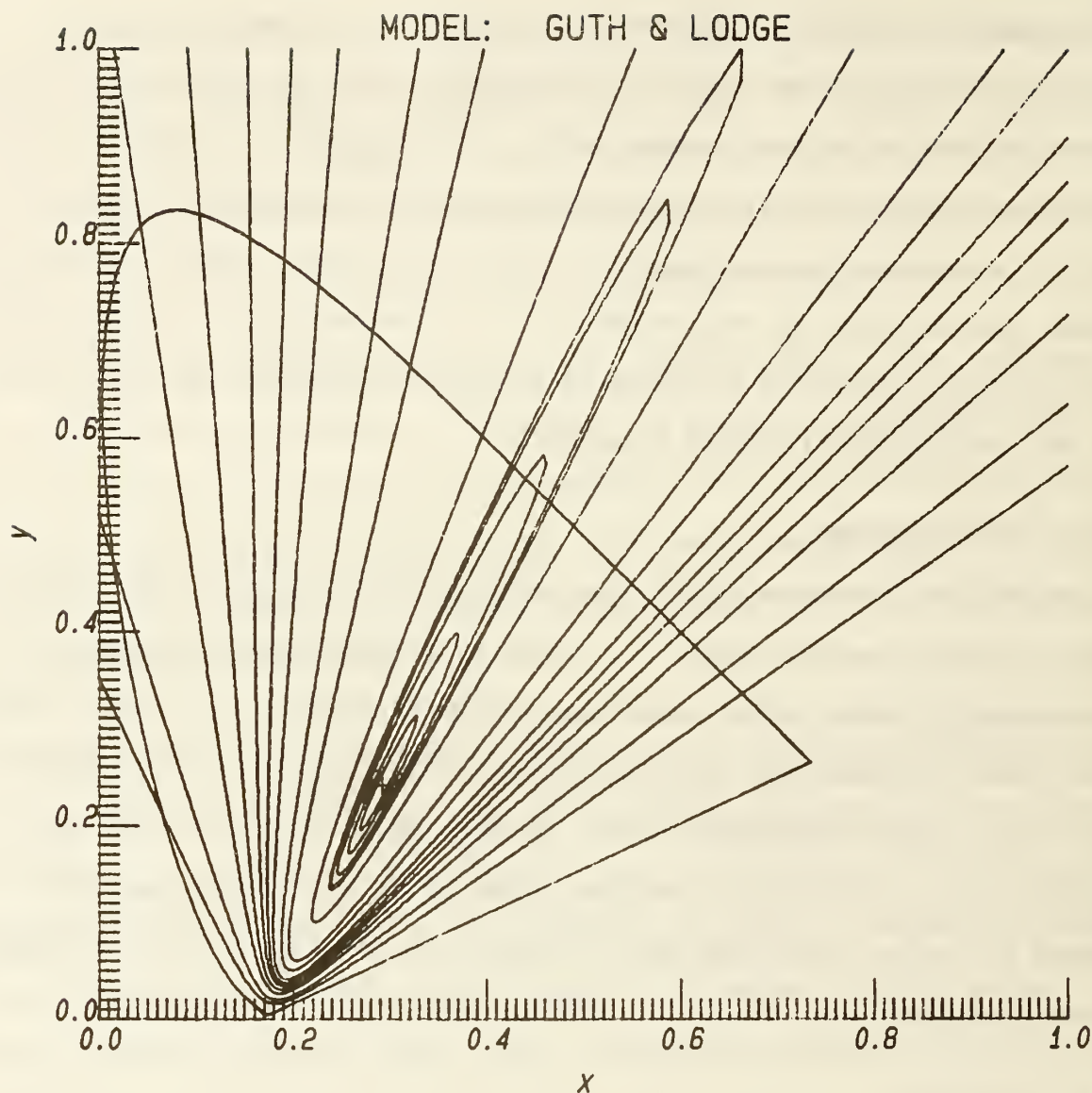


Figure 31. The loci of constant B/A (equivalent luminance over luminance) as predicted by the Guth and Lodge model. Because $p=2$ for this model, the curves are all conic sections (see text). At least some of the curves (those for low B/A values) are clearly ellipses (the asymmetrical clipping of some of the tips being an artifact of the plotting routine), whereas the curves for higher B/A appear to be hyperbolas. Fifteen ovals are shown. The B/A values, from inside (small) to outside (large) are: 1.0003, 1.0009, 1.002, 1.003, 1.0034, 1.01, 1.04, 1.06, 1.09, 1.11, 1.13, 1.16, 1.2, 1.3, and 1.4. The B/A value at 360 nm, the extreme shortwave end of the spectrum, is only 1.92 for this model. The plus sign (center of smallest oval) denotes the white point of the model, the only point at which $B/A=1$.

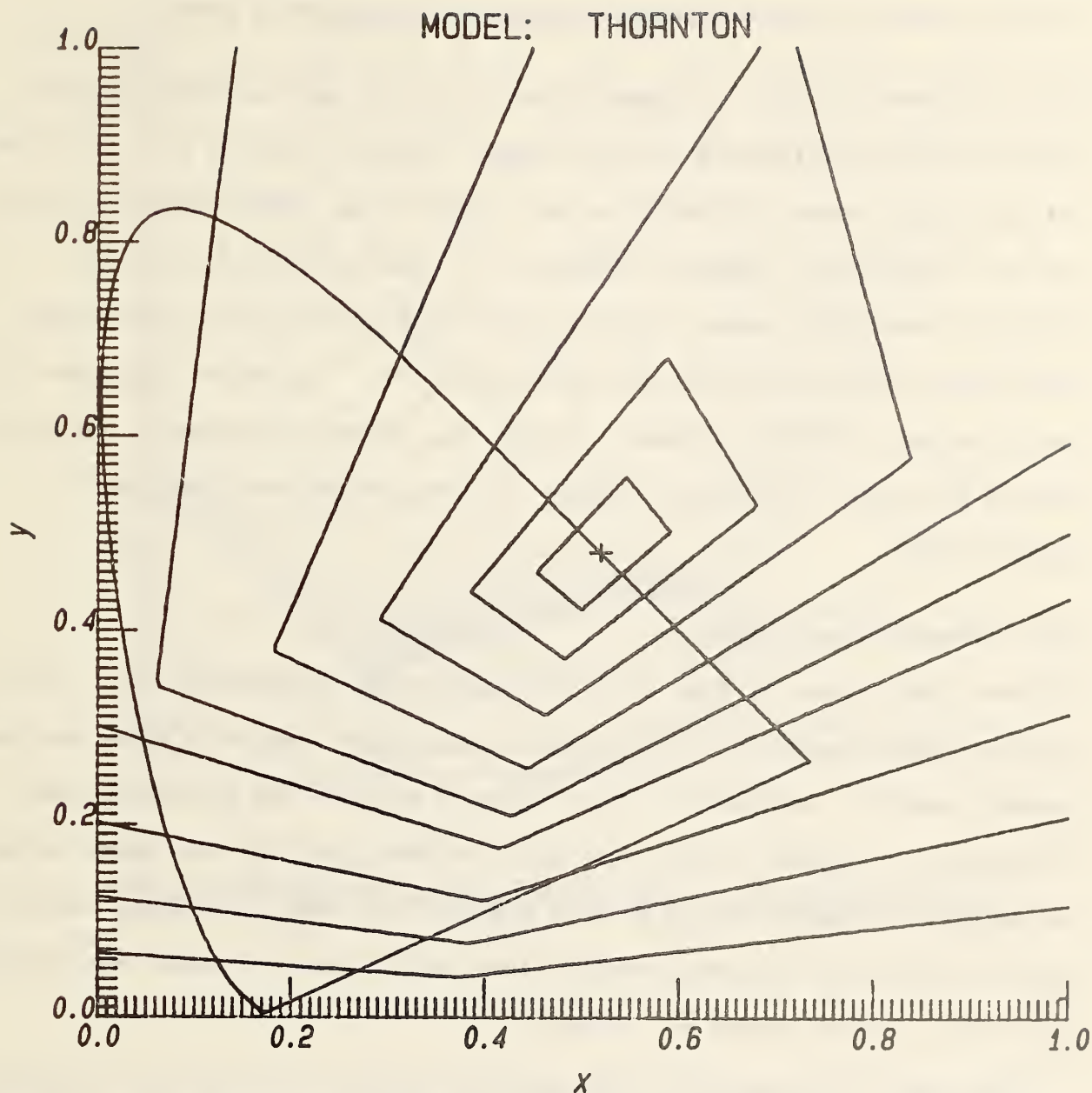


Figure 32. The loci of constant B/A (equivalent luminance over luminance) as predicted by the Thornton model. Because $p=1$ for this model, the loci are straight-sided quadrilaterals (see text), with sharp corners. Ten "ovals" are shown. The B/A values, from inside (small) to outside (large) are: 1.08, 1.17, 1.31, 1.5, 1.75, 2.0, 2.7, 4.0, 7.0, and 35.0. The B/A value at 360 nm, the extreme shortwave end of the spectrum, is a probably unrealistic 67.6 for this model. The plus sign denotes the "white" point of the model, the only point at which $B/A=1$.

plotting routine. In actuality, the intersections are sharp points.

Visual alignment in Fig. 32 suggests that all of the upper and lower corner points of the quadrilaterals lie on a single straight line, and that all of the left and right corners similarly lie on a single line. This appearance has not yet been confirmed by algebraic analysis. If there are indeed these two principal axes in the space, it is an interesting question as to what visual significance these directions have within the model. Comparison with Figs. 24 and 28 suggests that the alignment axes in Fig. 32 may correspond to the lines $T=0$ and $D=0$ within the Thornton model. This has not yet been confirmed algebraically.

6.8.6 Howett, Best-Fitting

The most novel aspect of Fig. 33 is the shape of the constant- B/A loci. They are four-sided figures with very sharp corners (cusps), and with sides that are concave outward. Comparison of Figs. 31 to 33 suggests the possibility that all models with powers greater than unity may have loci that are convex outward; all models with powers less than unity may have loci that are concave outward; and we do know that all models with a power exactly equal to unity have straight-sided loci (neither convex nor concave).

As with the Thornton quadrilaterals of Fig. 32, the "diamonds" of Fig. 33 seem to have their sharp corners aligned along the $T=0$ and $D=0$ lines of the model (see Figs. 25 and 29).

There is one extraordinary feature of Fig. 33. The lowest locus shown is close to the alychne ($A=0$) of the model (see Fig. 25). Because of division by zero, the constant- B/A locus coinciding with the alychne has a B/A value of infinity.

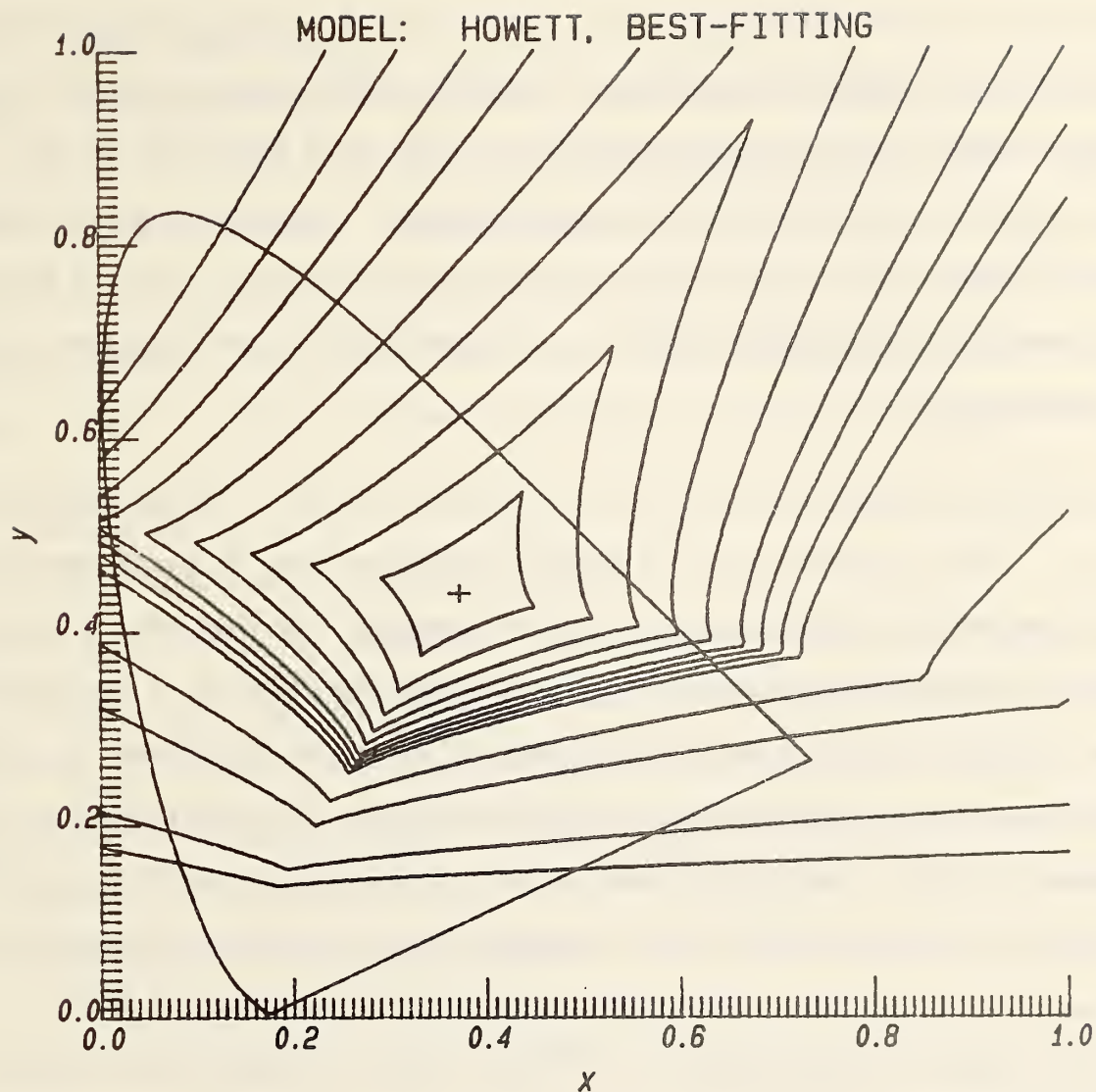


Figure 33. The loci of constant B/A (equivalent luminance over luminance) as predicted by the best-fitting Howett model. The equations for the "ovals" can be obtained by substituting the values from Eqs. (8)-(8a) into Eq. (26), or Eq. (31). Thirteen "ovals" are shown. The B/A values, from inside (small) to outside (large) are: 1.14, 1.23, 1.30, 1.36, 1.41, 1.46, 1.49, 1.52, 1.55, 1.75, 2.0, 3.0, and 4.0. A little below the line for $B/A=4.0$ lies the alychne ($A=0$), along which this model has infinite values of B/A . Below that, B/A becomes finite again, and reaches an absolute value of 3.17 at 360 nm, the extreme shortwave end of the spectrum. (The issue of whether the values below the alychne should conventionally be regarded as positive or negative is discussed in the text.) The plus sign denotes the white point of the model, the only point at which $B/A=1$.

Below that, the B/A ratio drops again to finite values. The sign of B/A below the alychne is a matter of convention. Because of the absolute values in its definition [Eq. (6)], B is always positive. Thus, from this point of view, when A goes negative, the ratio B/A also goes negative. As the alychne is crossed moving downward, the ratio switches from positive infinity to negative infinity, and thereafter rises (algebraically) to a modest -3.17 at the shortwave corner of the spectrum.

An alternative viewpoint is that it is reasonable always to give B the sign of A ; i.e., if the luminance (A) of a light is negative, then it seems appropriate for its equivalent luminance (B) to also be negative. By that interpretation, the ratio B/A is always positive, and, after reaching infinity on the alychne, drops back down to successively lower positive values as y continues to decrease. The B/A ratio at the shortwave end of the spectrum is, by this reasoning, best regarded as $+3.17$. The author tends to favor this latter convention, because it preserves the previously derived theorem that the ratio B/A can never be less than unity. If the first convention is adopted instead, then B/A values can only be greater than $+1$ or less than -1 , so that a gap in the permissible range of values then exists.

Of course, this entire discussion normally applies only to imaginary colors, lying outside the region of the chromaticity diagram that is of practical interest. In the case of the best-fitting Howett model, with which Fig. 33 deals, the critical crossover region involves physically realizable colors that can actually be observed in an experiment. This completely unrealistic behavior is still another aspect of why this model is not considered worth pursuing, in comparison with the restricted Howett model.

6.8.7 Howett, Restricted

Except for the bizarre behavior caused by the alychne running through the region of real colors in Fig. 33, the restricted Howett model illustrated in Fig. 34 has very similar features. Cusped diamonds may seem even harder to believe than Thornton's quadrilaterals, but the excellent fit of this model to the Sanders-Wyszecki data suggests that only a model that implies some such extreme shape—possibly slightly rounded—could agree well with that body of data.

6.9 ADDITIVITY

6.9.1 The Significance of Nonadditivity

Luminance is calculated, except for a possible constant factor, by weighting the spectral radiance of a light, wavelength by wavelength, by the $\bar{y}(\lambda)$ [$=V(\lambda)$] function, and summing these products across the entire visible spectrum. As a consequence of this quantitative definition, luminance is additive; that is, the luminance of the mixture of any two (or more) lights is equal to the sum of the luminances of the component lights.

Is equivalent luminance similarly additive? Since the definition of equivalent luminance is the luminance of an equally bright white light, an operational test of this question would run as follows: given, for example, a red light and a green light, match the red light in brightness by adjusting the luminance of a reference white light; similarly match the green light; and finally, match the mixture of the red and green lights. Is the luminance of the white light that matches the red-green mixture in brightness equal to the sum of the white luminances that match the red and green lights separately in brightness? As is now well known, the answer to the question is no. For most observers, the white luminance that matches the red-green mixture in brightness is significantly

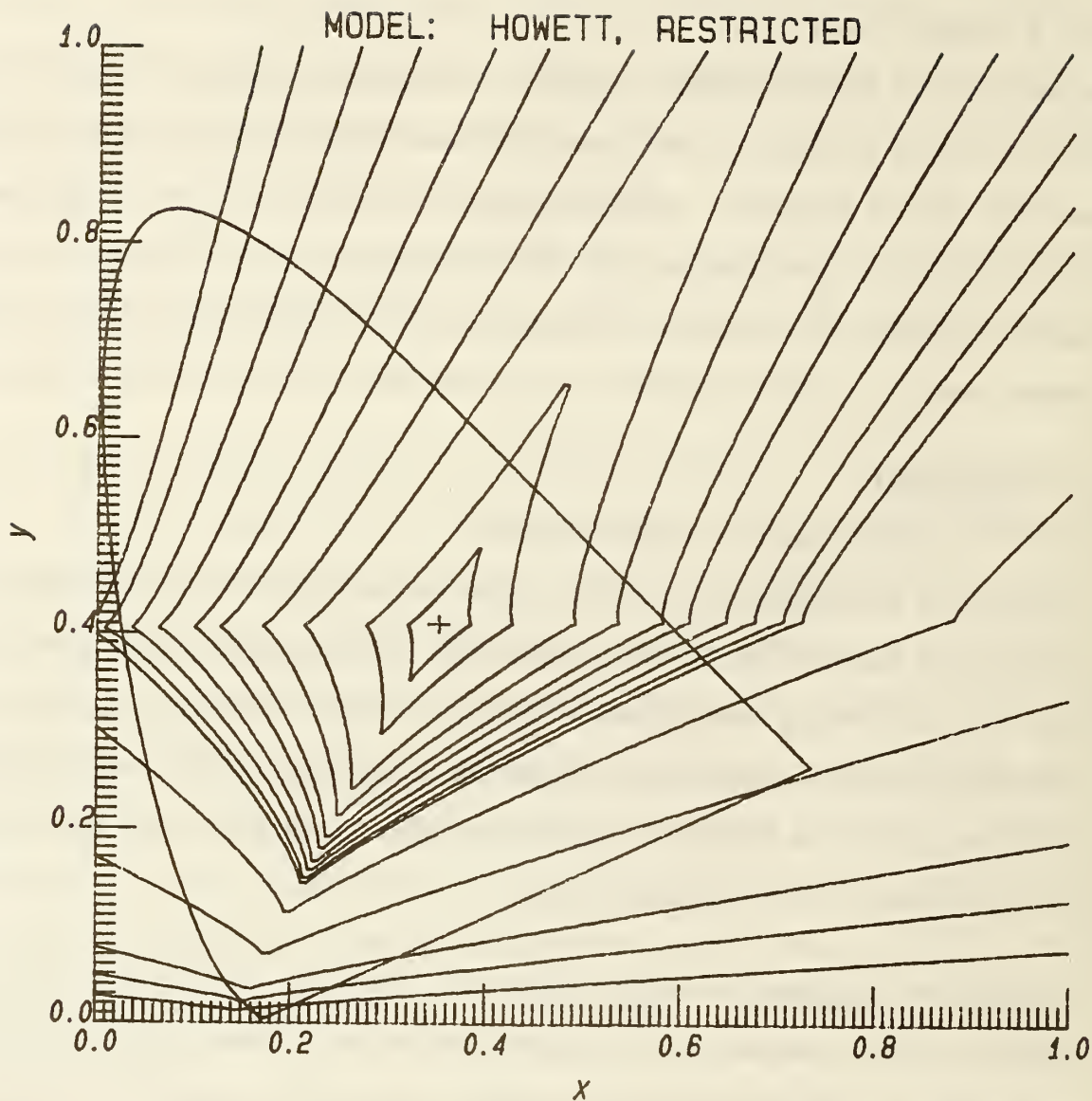


Figure 34. The loci of constant B/A (equivalent luminance over luminance) as predicted by the restricted Howett model. The equations for the "ovals" can be obtained by substituting the values from Eqs. (9)-(9a) into Eq. (26), or Eq. (31). Fifteen "ovals" are shown. The B/A values, from inside (small) to outside (large) are: 1.06, 1.12, 1.20, 1.25, 1.30, 1.33, 1.37, 1.40, 1.43, 1.45, 1.6, 2.0, 3.0, 4.0, and 6.0. The B/A value at 360 nm, the extreme shortwave end of the spectrum, is 15.9 for this model, a high but not ridiculous value. The plus sign denotes the white point of the model, the only point at which $B/A=1$.

less than the sum of the white luminances that brightness-match the red and green lights separately. Guth refers to this result as subadditivity. Where the equivalent luminance of a mixture exceeds the sum of the equivalent luminances of the component lights, Guth applies the term superadditivity. [These terms are widely used in the literature, and will be used here, but some authors (Kaiser and Wyszecki, 1978) prefer the expressions "additivity failure of the cancellation type," and "additivity failure of the enhancement type," respectively.] It is now well established that superadditivity does occur for some combinations of lights, although subadditivity is more common and more quantitatively extreme (Kaiser and Wyszecki, 1978; Burns et al., 1982).

The frequent discrepancy between equality of luminance and equality of brightness, for lights of different chromaticities, provides one of the two great challenges to color-vision models; the other is provided by the nonadditivity phenomenon. In developing his model, Guth confronted both of these challenges. His assumption that the brightness-perception center of the visual system is fed not only by the luminance channel, but also by the chromatic channels, accounts in a qualitative way for the imperfect correlation between brightness and luminance. His further assumption (not originated by him, of course) that the chromatic channels of the eye are opponent in nature allowed him to account qualitatively for nonadditivity. It is worth observing that a nonopponent (Young-Helmholtz) color vision model that assumes that one or more of the chromatic channels feeds into the brightness-perception center, together with a composite luminance channel, would also predict some discrepancy between luminance and brightness, but would entail perfect additivity. Thus, the two problems are closely related, but not logically equivalent.

6.9.2 Vectorial Considerations

Figure 35 is a sketch of the well-known principle of vector addition, shown in two dimensions. In the case shown, with a large angle between $\bar{Q}$ and $\bar{R}$, the length of the sum vector $\bar{Q}+\bar{R}$ is considerably less than the sum of the lengths of $\bar{Q}$ and $\bar{R}$. When the angle between $\bar{Q}$ and $\bar{R}$ is small, the length of $\bar{Q}+\bar{R}$ is only slightly less than the sum of the lengths of $\bar{Q}$ and $\bar{R}$. When the angle is zero--i.e., when $\bar{Q}$ and $\bar{R}$ are in the same direction--then the length of the sum vector is exactly equal to the sum of the lengths of the component vectors.

The review of the basic concepts of Guth's vector model given in section 1.1 indicates that, for him, $\bar{Q}$ and $\bar{R}$ would represent colors and $\bar{Q}+\bar{R}$ would represent the mixture color formed by superimposing the lights represented by $\bar{Q}$ and $\bar{R}$. Guth's space is three-dimensional, the axes being the quantities A, T, and D of Eqs. (1) and (2). The lengths of the vectors are proportional to the equivalent luminances (not the brightnesses) of the corresponding colors. It was indicated in the preceding subsection (6.9.1) that a simple index of additivity, as illustrated in Fig. 35, is the length of $\bar{Q}+\bar{R}$ minus the sum of the lengths of $\bar{Q}$ and $\bar{R}$. When this quantity is negative, subadditivity is indicated; when it is positive, superadditivity is indicated; and when it is zero, exact additivity is indicated. Because Guth's space has a euclidean metric [Eq. (2)], there is no arrangement of the vectors $\bar{Q}$ and $\bar{R}$ (Fig. 35) that can ever yield superadditivity.

6.9.3 Definition of a Generalized Nonadditivity Index

In terms of the generalized linear model represented by Eqs. (5) and (6), let us denote the length (norm) of a vector $\bar{C}$ by $|\bar{C}|^{(p)}$, the superscript (p) indicating the power in Eq. (6). Then we can define a generalized additivity--or really

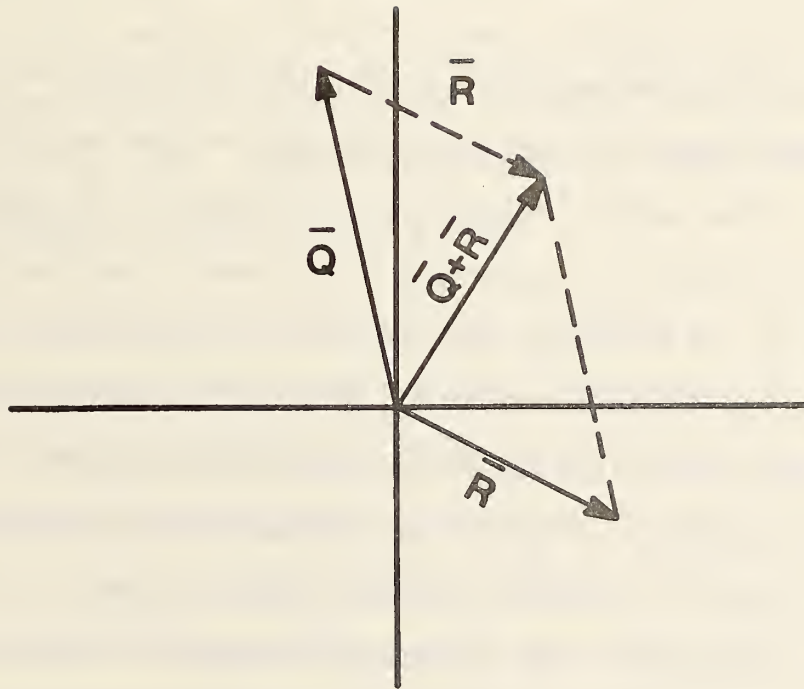


Figure 35. The principle of vector addition. $\vec{Q}$ and $\vec{R}$ (solid line) are vectors radiating from the origin. We translate $\vec{R}$ parallel to itself, until the tail of $\vec{R}$ (dashed) is touching the head (arrow) of $\vec{Q}$. Then the vector directed from the origin to the head of the translated $\vec{R}$ is the sum vector $\vec{Q}+\vec{R}$.

nonadditivity--index, applying for any value of p , as follows. For any colors C_i and C_j , and the mixed color C_{i+j} , the index is defined as

$$P^{(p)}(C_i, C_j) = \frac{100 \left[|C_{i+j}|^{(p)} - \left\{ |C_i|^{(p)} + |C_j|^{(p)} \right\} \right]}{|C_i|^{(p)} + |C_j|^{(p)}} . \quad (34)$$

By dividing the crude index mentioned earlier (6.9.2) by the sum of the vector norms, the $\underline{P}$ index of Eq. (34) becomes independent of the absolute sizes of the vectors; i.e., of the absolute luminances of the colors. The factor 100 converts the dimensionless index to a percentage, rather than a fractional basis.

Thus, a $\underline{P}$ value of -10 indicates that the norm of the sum vector (the equivalent luminance of the mixed color) falls short by 10 percent of equaling the sum of the separate vector norms (the sum of the equivalent luminances of the separate colors). This corresponds to what we can now describe as 10 percent subadditivity. Similarly, a $\underline{P}$ index of +5 denotes 5 percent superadditivity. A $\underline{P}$ value of 0 denotes perfect additivity. For the sake of precision of thought, $\underline{P}$ should not be described as an additivity index, but as a nonadditivity index; the bigger the absolute value of $\underline{P}$, the greater the degree of nonadditivity. For purposes of recalling the significance of the sign of the number, $\underline{P}$ can be thought of even more specifically as an index of percent superadditivity; but, in any case, it seems natural to associate "plus" with "super" and "minus" with "sub."

6.9.4 The Relationship of the Power, p , to Nonadditivity

It was pointed out in section 6.9.2 that when $p=2$ in Eq. (6) (the euclidean metric), a prediction of superadditivity is impossible. This is a significant failure of the Guth model and of the Ingling model. Is there any value of p

for which superadditivity can be predicted? In an attempt to find out how p influences the possible values of the P index, a Monte Carlo computer simulation was run. For simplicity, two-dimensional vectors were studied. The two components of each of two vectors were chosen by a random number generator that produced a uniform distribution of values ranging from -2 to $+2$. (This range was chosen in case values greater than unity in absolute value behave differently than values less than unity in absolute value.) The degree of superadditivity resulting from the addition of these randomly selected vectors was calculated in accordance with Eq. (34). Several hundred pairs of such randomly chosen vectors were examined for each value of p used.

The results of the Monte Carlo runs were quite unequivocal. For all p values exceeding unity, by even a very small increment, superadditivity never occurs. In other words, for $p > 1$, subadditivity always occurs (except for the vanishingly rare cases when the two vectors happen to lie along the same direction, or one of them is the zero vector, in which cases additivity occurs for any p). For $p = 1$, subadditivity is still the most common result, but perfect additivity occurs with significant frequency (about 25 percent) for vectors not lying along the same direction. It is only for p exactly equal to 1 that this phenomenon is observed. Superadditivity does not occur for $p = 1$. For p values less than unity, by even a very small decrement, subadditivity continues to be the most common condition, but superadditivity occurs about 25 to 30 percent of the time. Although it was obvious that the percentages might well vary with the dimensionality of the vector space, it seemed highly likely that the present author's models, with their powers in the neighborhood of 0.82, would be able to predict superadditivity in some cases. Whether these cases would correspond to the mixtures of real colors remained to be determined.

6.9.5 Actual Predictions: Pairs of Monochromatic Lights

Using the restricted model embodied in Eqs. (9)-(9a), additivity calculations were carried out, using the $\underline{P}$ index of Eq. (34), for mixtures of monochromatic lights. A complete matrix of combinations of wavelengths at 10-nm intervals over the visible spectrum was computed, for various luminance ratios of wavelength 1 (λ_1) to wavelength 2 (λ_2). The result for a 10:1 luminance ratio is shown in Fig. 36. The model of Eqs. (9)-(9a) can indeed predict superadditivity, and subadditivity as well. The surface depicted in the figure was somewhat smoothed by the plotting routine, but the principal features shown are accurate. The greatest degree of superadditivity is predicted by the model for long wavelengths combined with short wavelengths in a 10:1 ratio of the former to the latter (λ_1/λ_2). The greatest degree of subadditivity is predicted for middle wavelengths of λ_1 combined with either short or long wavelengths of λ_2 . Nearly perfect additivity is predicted for short wavelengths of λ_1 , regardless of the wavelength λ_2 . The particular luminance ratio of 10:1 was chosen for Fig. 36 to illustrate marked degrees of both sub- and superadditivity. Much larger and much smaller ratios of λ_1 to λ_2 yield predictions clustering more closely around the $P=0$ (additivity) plane.

6.9.6 Actual Predictions: Lines of Constant Dominant Wavelength

In an important paper combining theory and experiment, Burns et al. (1982) examined color mixture along lines of constant dominant wavelength. Such a line is traced out on the chromaticity diagram by combining a white with a monochromatic light in varying proportions, ranging from 0 (pure white light) to 1 (pure spectral light). The departures from additivity for the various mixtures along these lines were measured experimentally. Surprisingly, it

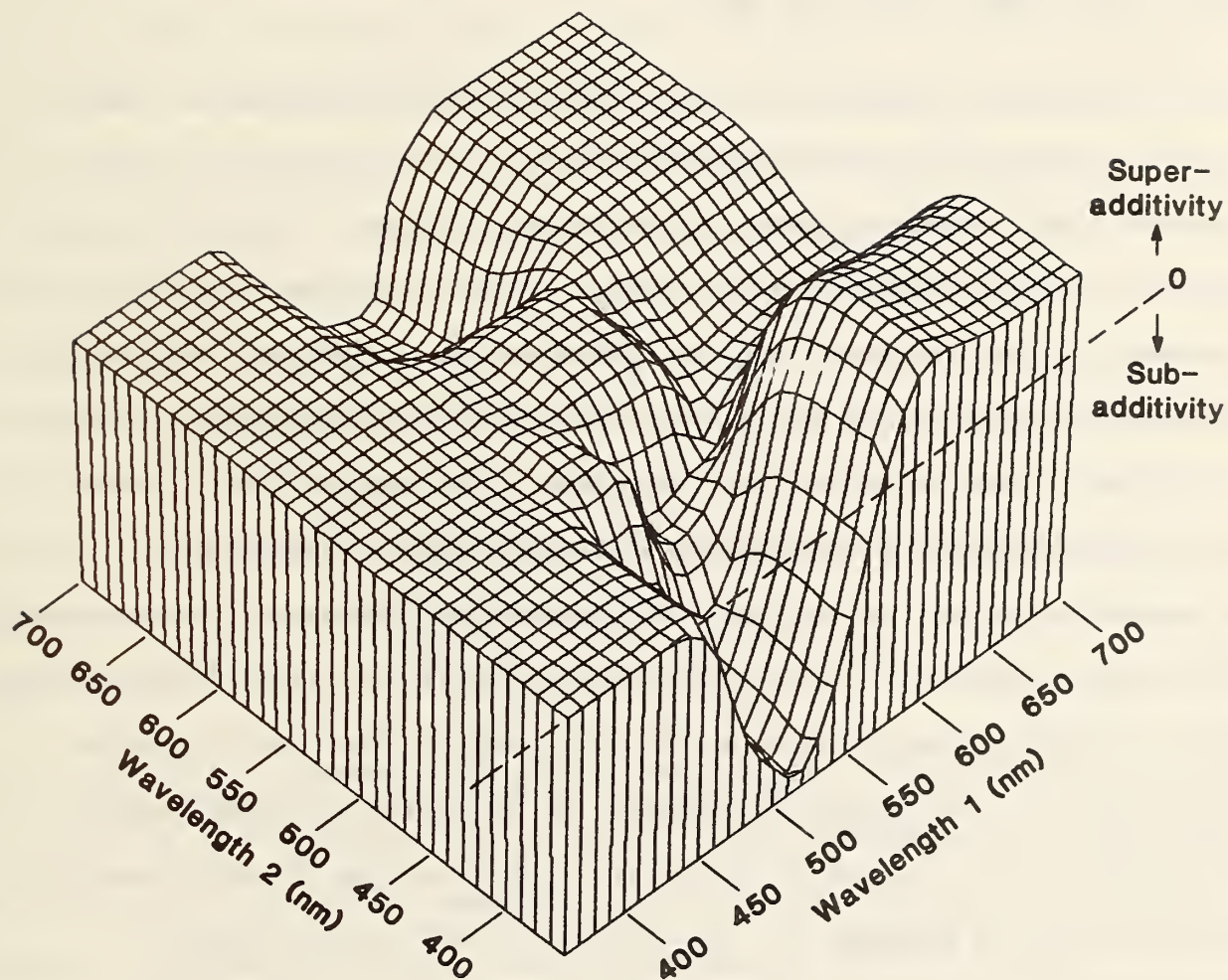


Figure 36. Predictions by the restricted Howett model [Eqs. (9)-(9a)] of the superadditivity index, $\underline{P}$ [Eq. (34)], for mixtures of two monochromatic lights. The plot shows $\underline{P}$ as a function of two wavelengths, presented in a mixture in the luminance ratio of 10:1 of λ_1 to λ_2 .

was found that for some dominant wavelengths, there was superadditivity at low purities (near white), with a switch over to subadditivity at high purities (near the spectrum locus). Of course, in such cases, there was one intermediate purity at which exact additivity held.

In their mathematical analysis of various types of models, Burns et al. (1982) appeared to conclude that no simple linear model, and not even some nonlinear models that they considered, could account for such results. However, careful inspection of the paper reveals that the generalized linear model [Eqs. (5)-(6)] for values of p less than unity was not examined by those authors. This omission was in all likelihood not an oversight; within the class of Minkowski L_p metrics [i.e., Eq. (6)], the range $p < 1$ is not considered to represent valid distance functions. L_p for any $p < 1$ is sometimes referred to as a "pseudometric." A true metric must satisfy certain basic mathematical conditions, and one of them is the triangle inequality. This condition requires that no side of any triangle can have a length greater than the sum of the lengths of the other two sides. In the light of all the preceding discussion of additivity and nonadditivity, within section 6.9, it should be clear that violation of the triangle inequality is precisely what is required if a quasi-metric is to be able to predict superadditivity. In fact, superadditivity is nothing more nor less than violation of the triangle inequality. Any true metric, which satisfies the triangle inequality, can never predict superadditivity.

The use of a p value less than unity in the model proposed in this report [Eqs. (9)-(9a)] means that the space corresponding to the model can be expected to exhibit strange or even bizarre behavior. It must be realized, however, that the facts of brightness perception--the data that the model is trying to predict

--are themselves strange or bizarre, and so far no predictions have been generated from the model that are absurd on their face.

The final additivity test that was carried out was to determine whether the restricted model of this paper [Eqs. (9)-(9a)] can predict the finding by Burns et al. that a crossover from one direction of nonadditivity to the other can occur along a line of constant dominant wavelength. Perhaps surprisingly, it turns out that it can. Figure 37 summarizes the results of these computations. What is plotted is the nonadditivity or superadditivity index, $\underline{P}$, of Eq. (34), as a function of dominant wavelength and colorimetric purity (the latter being the ratio of the monochromatic luminance to the sum of the white and monochromatic luminances). Each slice parallel to the purity axis shows the predictions for a particular dominant wavelength. Each slice represents a two-dimensional graph showing the superadditivity index $\underline{P}$ as a function of purity, for that dominant wavelength.

As is shown by the frontmost slice, for a wavelength of 700 nm, the prediction is that there is subadditivity at low purity values, additivity at a colorimetric purity of about 0.2, and then ^{super}~~sub~~additivity for purities beyond 0.2. Perfect additivity is predicted for the specific purities of 0 and 1.0, for all wavelengths, because in those two cases there is only one light in the mixture, and the general linear model [Eqs. (5)-(6)] for any $\underline{p}$ always predicts exact additivity when any light is added to itself or to nothing.

For short dominant wavelengths, particularly strong superadditivity is predicted at low purity values, and here there is no crossover; there is superadditivity for all but the terminal purities. Thus the qualitative pattern found by Burns et al. (1982)--crossover of the type of nonadditivity at some dominant wavelengths

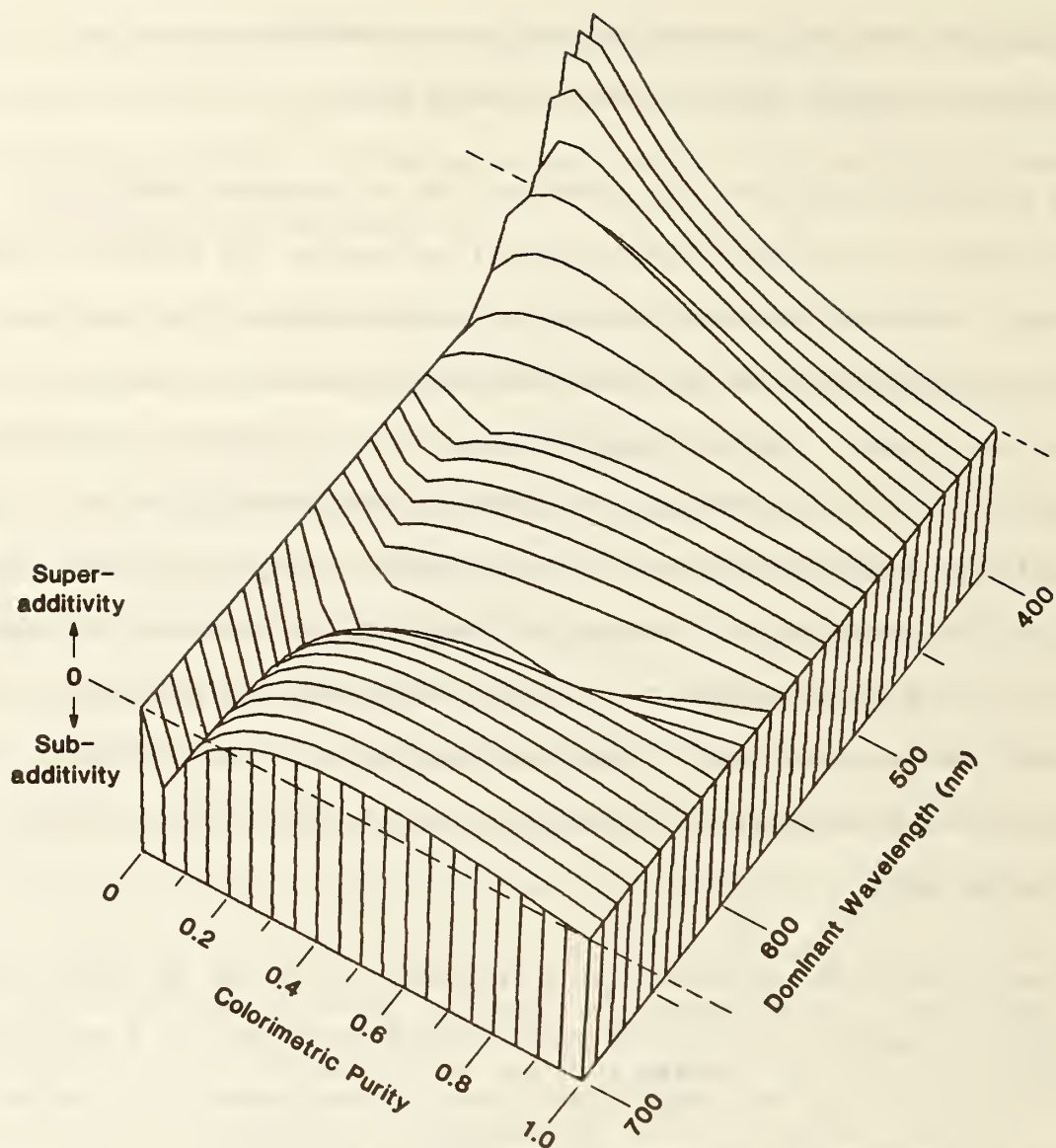


Figure 37. Predictions by the restricted Howett model [Eqs. (9)-(9a)] of the superadditivity index, $\underline{P}$ [Eq. (34)], for mixtures of white with monochromatic lights of different wavelengths. Each vertical section parallel to the purity axis shows the variation of predicted $\underline{P}$ with colorimetric purity (i.e., mixtures of the white and monochromatic lights in various luminance proportions). The chromaticity of the white used for the predictions was that of the actual white source used by Burns et al. (1982).

and nonadditivity of a single sign at other wavelengths--is predicted by the proposed model. Unfortunately, the details of the predictions differ sharply from the Burns et al. experimental data. For example, the predicted crossovers shown in Fig. 37 are from subadditivity at low purities to superadditivity at high purities, whereas the Burns et al. data exhibit the reverse direction of crossover. In their paper, Burns et al. used an indicator of nonadditivity different from the $\underline{P}$ index [Eq. (34)] used here. Their index, as well as $\underline{P}$, was calculated, and the two measures agree in their qualitative pattern. Thus the discrepancies between the predictions illustrated in Fig. 37 and the Burns et al. data must be due to a cause other than the different nonadditivity indexes used.

Kaiser and Wysecki (1978) showed that additivity information can be derived from the Sanders-Wysecki (1964) experimental data used as the basis for the model optimizations described in the present report. They presented only sketchy results, enough to show that both subadditivity and superadditivity are implicitly present in the Sanders-Wysecki data. Since the model proposed in this report [Eqs. (9)-(9a)] predicts the Sanders-Wysecki data so closely (see Fig. 2), it is reasonable to assume that a full additivity analysis of the Sanders-Wysecki data would lead to a pattern closely resembling Fig. 37. If that is true, then we have to confront the question of why the Sanders-Wysecki and Burns et al. data sets disagree with each other so much. Such disagreement seems to prevail between nearly any two studies in the literature. Kaiser and Wysecki (1978) seem to feel that a substantial part of the problem is the use of different groups of observers in the various experiments. That may be a factor, but it is also true that virtually every experiment in brightness perception is done with a different set of viewing conditions, experimental

procedures, and instructions to the observers. A systematic study of all these variables to pin down the sources of the disturbingly high inter-experiment variability seems to be a pressing need.

It has been demonstrated here that a linear model with a power less than unity can generate additivity predictions with the kinds of complex properties that characterize actual experiments. There seems reason to hope that if the general linear model [Eqs. (5)-(6)] were to be re-optimized to fit the Burns et al. data, a good fit might be possible, including reasonably detailed prediction of the complex additivity findings. One expectation that seems close to a certainty is that the value of p in the best-fitting Eq. (6) would be significantly less than unity.

7. SUMMARY AND CONCLUSIONS

The present author believes that averaging heterochromatic brightness-matching data from experiments carried out under different viewing conditions tends to obscure--rather than sharpen--the quantitative relationships underlying the behavioral data. As a result, this study represents an attempt to account mathematically for the data of a single, extensive experiment (Sanders and Wyszecki, 1964).

The essence of the approach was to formulate a generalized linear model of the opponent-colors type, and then optimize the constants of the model for best fit to the brightness-matching data. On the assumption that the optimization succeeded in locating the true global minimum of the sum of the squared prediction errors (and, unfortunately, it is not certain that this assumption was unequivocally fulfilled), it follows that the resultant model must necessarily fit the data better than any other linear, opponent-colors model can. A crucial

question is: how good is this best fit? That is, does the model predict the data well enough, in an absolute sense, that it would be considered a practical tool for calculating brightness matches among lights of different colors? The answer is yes: the predictions of the derived model correlate over 0.98 with the matching data.

The next crucial question is: how well would this formula generalize to predicting other sets of brightness-matching data? This question has not been answered here. It is certain that this one formula cannot accurately predict the data of all the experiments in the literature, since the various studies disagree sharply with each other in a quantitative--but not qualitative--sense. Such viewing parameters as the sizes of the fields being compared in brightness, the separation between them, the luminance level of the lights, the presentation time if the viewing is flashed, and the chromaticity and luminance of the background, all are expected to influence equal-brightness judgments between lights of different colors. Ultimately, a formula for predicting equivalent luminance (the luminance of an equally bright reference white light) should incorporate explicit adjustment terms for each of these factors. Thus, a great deal of experimental, mathematical, and theoretical work clearly remains to be done before such a highly refined formula can be derived.

The formula for equivalent luminance proposed by the present author [Eqs. (5), (6), (9), and (9a)] is recommended for testing in "real-life" situations. It is of particularly simple form: it represents a linear opponent-colors model with a combination rule (metric) different from previous proposals. Although evidence is accumulating that some sort of nonlinearity needs to be introduced in an equivalent-luminance formula in order to account for certain color-vision

data, it is of interest to explore how well the best possible linear formula can do. That is why an optimization procedure was used in the present study.

Even if the new formula were to be regarded as "mere curve-fitting," it would be desirable to see how useful it proves to be in practical lighting work involving lights with significant coloration. If it proves to be more realistic than other formulas currently available, it can be used until something better comes along. However, the present study has shown that the model of color vision implied by the formula is capable of predicting at least the gross forms of some of the classical color-vision functions, so that something more than "mere curve-fitting" may be represented by the model. It is also an encouraging sign that the channel responses that resulted from a completely unrestricted optimization of the generalized linear model [Eqs. (5)-(6)] turned out to consist of a nearly all-positive channel resembling the $\bar{y}(\lambda)$ function, plus two opponent channels corresponding more or less to red-green and blue-yellow responses.

Moreover, when a new optimization was done with one channel restricted to being all-positive everywhere within the visible spectrum, that channel turned out to be essentially equal to $\bar{y}$, and again the other channels were opponent in form and more or less identifiable as red-green and blue-yellow. Thus, heterochromatic brightness-matching data seem to directly imply that processes resembling opponent-colors theory are actually operative in the eye. Some further exploration of linear versions of that theory may still prove informative, although nonlinear versions may take over as further data are published.

The present author's formulation of the generalized linear model involves defining equivalent luminance as the p th root of the sum of the p th powers

of the three channel responses. For Guth and for Ingling, $p=2$, and for Thornton, $p=1$. The optimizations performed as part of the present analysis show that a power in the neighborhood of 0.82 is necessary to account for at least the particular body of heterochromatic brightness data that was studied. This is one of the most distinctive findings of the present study, and its implications would seem to deserve further exploration. The analysis of the known nonadditivity of equivalent luminances carried out in this report has already shown that no linear model with a p value exceeding unity can predict superadditivity at all; and that the author's model, with its power less than unity, can predict the sort of complex pattern of mixed sub- and superadditivity that characterizes real brightness-matching data.

7.1 SUGGESTIONS FOR FURTHER WORK

A great deal of additional work in the general subject area covered by this report is desirable in order to clarify the many unsettled issues. It is clear that such further exploration needs to include both experimental and computational investigations. Three examples of each type are described in the following subsections.

7.1.1 Some Proposed Experimental Studies

7.1.1.1 Is Chromatic Brightness Visually Effective?

Strongly colored light (other than yellow) certainly looks brighter than whiter light, but can it help us to see better? An experiment to answer this question needs to concern itself with suprathreshold vision, since the eye's functioning at spatial or temporal threshold appears to be controlled by luminance, rather than brightness. If chromatic brightness does make a contribution to suprathreshold visual performance, a potential arises for saving energy by using

lower illuminances of somewhat chromatic light, without loss of visual function.

7.1.1.2 How Much Coloration in a Light Source Is Tolerable?

An important limitation on the use of strongly chromatic lights in everyday work environments is imposed by people's frequently observed intolerance for prolonged exposure to such lights. Presumably, the upset that people often feel when forced to work under substantially chromatic lights is associated with the progressive degradation of the color rendering of familiar objects as a light source departs increasingly from white. It would be very useful to know how saturated a light source, in every hue direction, is acceptable to the average room occupant during long-term exposure.

7.1.1.3 The Effect of Viewing Conditions on Equivalent Luminance

It is of crucial importance to explain the quantitative disagreement among the various experimental studies of heterochromatic brightness matching in the literature. Part of the variation almost surely arose because of the use of different viewing conditions by the various investigators. What is needed is a parametric investigation of the effects on brightness perception of separate and simultaneous variation of a small number of principal properties of the stimulus fields. A first experiment might examine the influences of the size and separation of the fields being compared, as well as the luminance level. Later studies could concern themselves with the relevant properties of the background field, particularly its chromaticity and luminance. A key goal is the inclusion of explicit functions of the more critical of these parameters as modifiers of the constants in color-vision models.

7.1.2 Some Proposed Computational Studies

7.1.2.1 Individual Observer Variation

Some of the published experiments on heterochromatic brightness matching include data for the individual observers. If an efficient optimization routine can be discovered, it would be of great interest to optimize the constants of the general linear model for each observer separately. It would then be possible to note which parameters of the model tend to remain relatively constant over different observers, and which ones tend to vary the most. The correlations among the parameters could also be examined. If enough correlation exists, an attempt could be made to characterize each observer by a small number of constants from which all the constants of the observer's best-fitting model could be generated.

7.1.2.2 Studies of Nonlinear Models

Various types of nonlinear models need to be investigated. It is important to determine how much better various color-vision functions can be predicted by those nonlinear models than by the best linear models. The statistics of nonlinear prediction might be applied in this connection, for significance testing. A determination should also be made as to what color-vision phenomena are correctly predicted by the nonlinear models, but not by any linear model.

7.1.2.3 Further Predictions of Color-Vision Functions by The Optimized Linear Model of This Report

Determination of how well the model embodied in Eqs. (9)-(9a) predicts color-vision functions not considered in this report should be informative. In addition, attempts should be made to improve the fit of some of the functions that have been considered here, by revising the equations postulated as connecting the model to the various color-vision functions.

8. REFERENCES

- Benzschawel, T.; and Guth, S. L. (1984). "ATDN: toward a uniform color space," Color Research and Application, Vol. 9, No. 3 (Fall), pp. 133-141.
- Burns, S. A.; Smith, V. C.; Pokorny, J.; and Elsner, A. E. (1982). "Brightness of equal-luminance lights," Journal of the Optical Society of America, Vol. 72, No. 9 (Sept.), pp. 1225-1231.
- CIE (1970). International Lighting Vocabulary, 3rd Edition, CIE Publication No. 17, Central Bureau of the CIE, Paris.
- CIE (1978). Light as a True Visual Quantity: Principles of Measurement, CIE Publication No. 41, Central Bureau of the CIE, Paris.
- Guth, S. L.; and Lodge, H. R. (1973). "Heterochromatic additivity, foveal spectral sensitivity, and a new color model," Journal of the Optical Society of America, Vol. 63, No. 4 (April), pp. 450-462.
- Guth, S. L.; Massof, R. W.; and Benzschawel, T. (1980). "Vector model for normal and dichromatic color vision," Journal of the Optical Society of America, Vol. 70, No. 2 (Feb.), pp. 197-212.
- Howett, G. L. (in press). "The coming redefinition of photometry." Scheduled for publication in Journal of the Illuminating Engineering Society, Vol. 15, No. 3, April 1986.
- Ingling, C. R., Jr. (1977). "The spectral sensitivity of the opponent-color channels," Vision Research, Vol. 17, No. 9, pp. 1083-1089.
- Ingling, C. R., Jr.; and Tsou, B. H.-P. (1977). "Orthogonal combination of the three visual channels," Vision Research, Vol. 17, No. 9, pp. 1075-1082.
- Judd, D. B. [unattributed] (1951). "Report of Secretariat of Technical Committee No. 7, Colorimetry and Artificial Daylight," CIE Proceedings of Twelfth Session, Stockholm, 1951, Vol. 1, Central Bureau of the CIE, New York, p. 7/11.
- Kaiser, P. K.; and Wyszecki, G. (1978). "Additivity failures in heterochromatic brightness matching," Color Research and Application, Vol. 3, No. 4 (Winter), pp. 177-182.
- Sanders, C. L.; and Wyszecki, G. (1964). "Correlate for brightness in terms of CIE color matching data," CIE Compte Rendu, Fifteenth Session, Vienna, 1963, Vol. B, CIE Publication No. 11, Central Bureau of the CIE, Paris, pp. 221-230.
- Thornton, W. A. (1973). "A system of photometry and colorimetry based directly on visual response," Journal of the Illuminating Engineering Society, Vol. 3, No. 1, pp. 99-111.
- Vos, J. J. (1978). "Colorimetric and photometric properties of a 2° fundamental observer," Color Research and Application, Vol. 3, No. 3 (Fall), pp. 125-128.

APPENDIX: BASIC COLORIMETRIC QUANTITIES AND NOTATION

CAPITAL LETTERS

Formal colorimetric systems are based on the fundamental notion that light of any color can be matched by a mixture of three fixed lights called primaries. The amounts of the three primaries needed to match any color are called the tristimulus values of that color.

For each of the primary lights, a particular luminance is defined, in effect, as unit amount of that primary. (The definition is usually indirect, and numerical luminance values for the units are rarely seen, as such.)

Tristimulus values are specified relative to these unit quantities of each primary. Thus, each tristimulus value is a ratio of one luminance to another, and is dimensionless.

Tristimulus values are conventionally denoted by capital letters corresponding to the designations of the primaries within the system. In the CIE 1931 standard colorimetric system, the tristimulus values are designated X, Y, and Z, corresponding to the amounts of the particular red, green, and blue primaries of that system, respectively. (The underlining here only indicates italic letters; see the Note on Notation immediately preceding section 1 of this report.)

In specifying the tristimulus values of a group of samples, the absolute sizes of the tristimulus values of each sample have significance relative to the sizes of the tristimulus values of the other samples. However, the overall absolute scale of tristimulus values is arbitrary. In the case of patches of light (the type of stimuli referred to in this paper), a common convention is to set the Y value of each stimulus equal to its luminance, in some convenient

units such as footlamberts or nits (cd/m^2). The X and Z values are adjusted proportionally. Thus, the relative luminances of any two samples can be compared by simply noting their respective Y values.

In opponent-colors systems, the quantities analogous to tristimulus values are also denoted by capital letters. Accordingly, following Guth (Guth and Lodge, 1973), this paper uses the notation A, T, and D for the amounts of stimulation of the luminance, red-green, and blue-yellow channels, respectively.

LOWER-CASE LETTERS

When the issue of luminance is secondary, and only color (in the popular sense) is of interest, it is only the proportions of the tristimulus values to each other that are relevant, and not their absolute size. Hence, the quantities

$$x = \frac{X}{X+Y+Z}, \quad y = \frac{Y}{X+Y+Z}, \quad z = \frac{Z}{X+Y+Z} \quad (\text{A1})$$

are defined, and referred to as chromaticity coordinates. From the definitions of x, y, and z given in Eq. (A1), it is easy to see that: (a) the chromaticity coordinates sum to unity; and (b) the tristimulus values are proportional to the chromaticity coordinates, the proportionality constant being $X+Y+Z$. If, as is common, stimuli are specified by x, y, and Y--two chromaticity coordinates and the luminance of the light--then the z coordinate is calculated as $1-x-y$; and the X and Z tristimulus values are retrieved by multiplying x and z, respectively, by the ratio Y/y . Because of property (a), only two of the chromaticity coordinates are independent. Therefore, the standard chromaticity diagram represents the basic luminance-independent aspects of color in the form of a plot of y against x, with no reference to z.

The CIE 1931 x,y chromaticity diagram has one conspicuous failing: its spacing is not uniform. A uniform [or uniform-chromaticity-scale (UCS)] diagram is one having the property that the distance between any two chromaticity points on the diagram is proportional (at least approximately) to the amount of difference perceived between the colors corresponding to the points. In 1960, the CIE recommended, for use as an approximately uniform chromaticity diagram, an alternate set of coordinates designated u and v. The u,v diagram is described mathematically as a projective transformation of the x,y diagram, which means that straight lines on the x,y diagram are transformed into straight lines on the u,v diagram, and vice versa. One of the most important properties of a chromaticity diagram is that all colors that can be formed by mixing two fixed colored lights in different proportions have chromaticities that lie along the straight line segment in the diagram joining the chromaticities of the two fixed colors. Consequently, the u,v diagram, despite its different spacing, represents exactly the same facts about color mixture as does the x,y diagram.

Although this report does not make any use of the opponent-colors analogues of chromaticity coordinates, if such quantities had been used, they would have been denoted by the lower-case letters a, t, and d.

LOWER CASE LETTERS WITH OVERBARS

The first experimental step in the development of a system of colorimetry is to determine the amounts of the three primaries needed to match any monochromatic (single-wavelength) light within the visible spectrum. (In practice, these data are measured at intervals of several nanometers--usually 5 or 10 nm--and then interpolated to 1-nm intervals.) These are the tristimulus values of the spectrum. The particular spectrum that is used is usually an "equal-energy"

spectrum: one produced by the emission of equal amounts of power (as measured by radiance, for example) within each small, constant wavelength band across the visible spectrum.

These spectral tristimulus values, which are referred to in older literature as "distribution coefficients," are denoted by lower-case letters with overbars, as: $\bar{x}(\lambda)$, $\bar{y}(\lambda)$, and $\bar{z}(\lambda)$. The (λ) [Greek lambda] indicates functional dependence on wavelength, and is redundant, since lower-case overbarred letters are used only in connection with the spectral tristimulus values. Therefore, the simple symbols $\bar{x}$, $\bar{y}$, $\bar{z}$ are frequently used without explicit inclusion of λ .

One of the most important properties of tristimulus values is that they are additive: that is, the tristimulus values of a compound stimulus obtained by superimposing several lights is equal to the sum of the tristimulus values of the separate, component lights. Consequently, the tristimulus values of any stimulus that has been analyzed into its component wavelengths--that is, whose spectral power distribution is known--can be obtained by adding together the tristimulus values of the spectrum, weighted in each wavelength band by the amount of energy (radiance) in that band in the spectrum of the stimulus. In short, the spectral tristimulus-value functions $\bar{x}(\lambda)$, $\bar{y}(\lambda)$, and $\bar{z}(\lambda)$ constitute standard weighting functions which, when used to weight the spectral power distribution of any light, yield, after totalling across the entire spectrum, the X,Y,Z tristimulus values of the light. That is how tristimulus values are usually actually calculated.

Still another way of describing the same arithmetic is to say that $\bar{x}(\lambda)$,

$\bar{y}(\lambda)$ and $\bar{z}(\lambda)$ represent the spectral sensitivities of three "receptor mechanisms" on which light is shone. The total response of the $\bar{x}$ mechanism when any light is "absorbed" in it is the $\bar{x}$ tristimulus value; and similarly for the other two "mechanisms." In the case of CIE $\bar{x}, \bar{y}, \bar{z}$, it is understood that the $\bar{x}$, $\bar{y}$, and $\bar{z}$ "mechanisms" are formalisms only, and do not represent, for example, the absorptions of the three types of color receptors in the actual human eye. Nevertheless, it can be shown that $\bar{x}(\lambda)$, $\bar{y}(\lambda)$, and $\bar{z}(\lambda)$ must be linear combinations of the spectral sensitivities (absorptances) of the actual eye mechanisms. The spectral tristimulus functions of any other colorimetric system which contains the same information as the CIE system must likewise be linear combinations of the photoreceptor sensitivity functions of the eye. The conclusion is that all (infinitely many) CIE-equivalent colorimetric systems can be expressed as linear transformations of each other. Accordingly, one may work in any system convenient to the task at hand, and then easily convert later to some fixed reference system, such as the CIE 1931 system. Because of the relationship, described above, of the CIE system to the actual color receptor mechanisms of the eye, it is assumed that any two light stimuli having equal tristimulus values will appear to be a perfect visual match to the standard human eye (which is supposed to represent the average of the entire population with normal color vision). Since the $\bar{x}, \bar{y}, \bar{z}$ functions determine what the calculated tristimulus values of any stimulus are, the functions are also sometimes referred to as the color-matching functions.

With regard to opponent-color systems, the analogues of spectral tristimulus values are also denoted by lower-case overbarred letters, as $\bar{a}$, $\bar{t}$, and $\bar{d}$; or, when the dependence of the functions on wavelength is to be emphasized, as $\bar{a}(\lambda)$, $\bar{t}(\lambda)$, and $\bar{d}(\lambda)$. These functions are thought of as the response sensitivities

of the three opponent-level mechanisms--luminance, red-green, and blue-yellow, respectively.

Capital letters with overbars have no accepted meaning in colorimetry. In this report, they have been used to symbolize vectors (see Note on Notation, p. xi).

U.S. DEPT. OF COMM. BIBLIOGRAPHIC DATA SHEET <i>(See instructions)</i>	1. PUBLICATION OR REPORT NO. NBSIR/85-3202	2. Performing Organ. Report No.	3. Publication Date FEBRUARY 1986
4. TITLE AND SUBTITLE Linear Opponent-Colors Model Optimized for Brightness Prediction			
5. AUTHOR(S) Gerald L. Howett			
6. PERFORMING ORGANIZATION (If joint or other than NBS, see instructions) NATIONAL BUREAU OF STANDARDS DEPARTMENT OF COMMERCE ~~WASHINGTON, D.C. 20534~~ Gaithersburg, MD 20899			7. Contract/Grant No. 8. Type of Report & Period Covered Final
9. SPONSORING ORGANIZATION NAME AND COMPLETE ADDRESS (Street, City, State, ZIP) National Bureau of Standards - Department of Commerce Gaithersburg, MD 20899 and U.S. Department of Energy Washington, D.C. 20585			
10. SUPPLEMENTARY NOTES ☐ Document describes a computer program; SF-185, FIPS Software Summary, is attached.			
11. ABSTRACT (A 200-word or less factual summary of most significant information. If document includes a significant bibliography or literature survey, mention it here) Formal multivariate optimization techniques were applied in an attempt to determine how well a linear, opponent-colors model of color vision could account for specific brightness-matching data. The data fitted were from a single experiment by Sanders and Wyszecki that matched an adjustable white light in brightness to each of a set of lights of 96 different colors and constant luminance. A generalized, linear, opponent-colors model was formulated, which included the models of Guth (and co-workers), Ingling (and co-workers), and Thornton as special cases. The model contained 10 parameters, including nine determining the spectral responses of the three opponent-level channels and one determining the rule for combining the outputs of the three channels to obtain an estimate of equivalent luminance (the luminance of an equally bright white light). Despite difficulties with the optimization procedure, a model was found that correlates better than 0.98 with the fitted data. The predictions of this model for various other color-vision functions were explored and compared with corresponding predictions of the Guth and Lodge model and the Thornton model. The new model's predictions of these functions are less than perfect, but surprisingly good considering that the model was optimized entirely on brightness data (the only restriction being that the luminance channel should not have any negative values). The model was shown to predict the sort of complex mixture of sub- and superadditivity that is present in actual data. Some new algebraic results concerning the "B/Y" or "B/L" (equivalent luminance over luminance) ovals on the chromaticity diagram were derived.			
12. KEY WORDS (Six to twelve entries; alphabetical order; capitalize only proper names; and separate key words by semicolons) additivity; brightness; color; equivalent luminance; Guth model; luminance; model, color vision; opponent colors; optimization, multivariate; primaries; white point			
13. AVAILABILITY ☒ Unlimited ☐ For Official Distribution. Do Not Release to NTIS ☐ Order From Superintendent of Documents, U.S. Government Printing Office, Washington, D.C. 20402. ☒ Order From National Technical Information Service (NTIS), Springfield, VA. 22161			14. NO. OF PRINTED PAGES 126 15. Price \$16.95

